

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 543.069 Seconds  
(without alignments)  
1596.226 Million cell updates/sec

Title: US-10-624-714-11  
Perfect score: 20  
Sequence: 1 cgaggccgcccacacaa 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residuiss

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 300 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_in.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pi.\*  
9: gb\_pi.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_mu.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pi.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_hgt\_hum.\*  
31: em\_hgt\_inv.\*  
32: em\_hgt\_inv.\*  
33: em\_hgt\_mus.\*  
34: em\_hgt\_pln.\*  
35: em\_hgt\_rod.\*  
36: em\_hgt\_mam.\*  
37: em\_hgt\_vrt.\*  
38: em\_sy.\*  
39: em\_hgt\_hum.\*  
40: em\_hgt\_mus.\*  
41: em\_hgtg\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	95.0	17131	1	AE007080 Mycobacte
2	19	95.0	306550	1	EX248342 Mycobacte
3	19	95.0	348247	15	EX842579 Mycobacte
C 4	17.4	87.0	1073	1	RMU22370 Rhizobium m
C 5	17.4	87.0	2253	9	AK098077 Homo sapi
C 6	17.4	87.0	10029	1	AE007196 Sinorhizo
C 7	17.4	87.0	10296	1	AE007297 Sinorhizo
C 8	17.4	87.0	10578	1	AE007314 Sinorhizo
9	17.4	87.0	10931	1	AE007285 Sinorhizo
10	17.4	87.0	12706	4	BTINXGENE
11	17.4	87.0	37476	9	HS316G12 Human DNA
C 12	17.4	87.0	86834	9	AF003626 Homo sapi
C 13	17.4	87.0	97834	2	AL354677 Homo sapi
14	17.4	87.0	105589	2	AC092269 Homo sapi
C 15	17.4	87.0	129679	9	AF187000 Homo sapi
16	17.4	87.0	154807	2	AC069493 Homo sapi
C 17	17.4	87.0	166847	9	AC008393 Homo sapi
18	17.4	87.0	175511	2	AC023924 Homo sapi
19	17.4	87.0	181070	2	AC142187 Rattus no
20	17.4	87.0	182765	2	AC124790 Homo sapi
21	17.4	87.0	186908	2	AC124791 Homo sapi
22	17.4	87.0	189333	1	RME603647 Rhizobium
C 23	17.4	87.0	189436	9	AF274573 Homo sapi
C 24	17.4	87.0	195563	9	AC100872 Homo sapi
C 25	17.4	87.0	197000	2	AC124859 Homo sapi
C 26	17.4	87.0	202437	2	AC023224 Homo sapi
27	17.4	87.0	260050	1	RME603643 Rhizobium
28	17.4	87.0	279011	9	AE006467 Homo sapi
29	17.4	87.0	286550	1	SMES91785 Sinorhizo
C 30	17.4	87.0	286550	1	SMES91785 Sinorhizo
C 31	17.4	87.0	299350	1	SMES91786 Sinorhizo
C 32	17.4	87.0	300000	1	SMES91784 Sinorhizo
33	17.4	87.0	333800	1	SMES91792 Sinorhizo
34	17	85.0	1867	9	AF013956 Homo sapi
C 35	17	85.0	88655	9	AC079135 Homo sapi
36	17	85.0	118404	4	AL773591 Pig DNA s
C 37	17	85.0	175127	2	AC016048 Homo sapi
C 38	17	85.0	189050	2	AC142566 Bos tauru
C 39	17	85.0	190050	1	AL646059 Ralstonia
C 40	17	85.0	199827	2	AC026373 Homo sapi
41	17	85.0	206960	2	AC128121 Rattus no
42	17	85.0	209751	9	AC087645 Homo sapi
C 43	17	85.0	214485	2	AC123416 Rattus no
C 44	17	85.0	215313	9	AC100791 Homo sapi
C 45	17	85.0	229426	2	AC010532 Homo sapi
C 46	17	85.0	246860	2	AC115225 Rattus no
C 47	17	85.0	268510	2	AC095840 Rattus no
48	17	85.0	320746	2	AC128373 Rattus no
C 49	16.8	84.0	442	9	HS4236689
C 50	16.8	84.0	993	6	AX523503 Sequence
C 51	16.8	84.0	1002	6	AX497801 Sequence
52	16.8	84.0	4588	5	AF380137 Takifugu
53	16.8	84.0	8851	1	PD034353 Paracoccus
54	16.8	84.0	9851	6	AX647145 Sequence
55	16.8	84.0	13860	1	AE001247 Treponema
56	16.8	84.0	22203	5	AY081135 Takifugu
57	16.8	84.0	25155	2	AC019899 Drosophila
58	16.8	84.0	35728	3	U41022 Caenorhabdi
C 59	16.8	84.0	37113	6	AX708626 Sequence
60	16.8	84.0	63964	2	AC126346 Homo sapi
C 61	16.8	84.0	65201	9	HS754E20A Human DNA
62	16.8	84.0	122732	2	HS436C18 Homo sapi
C 63	16.8	84.0	138229	2	HSAC001236 Homo sapi
64	16.8	84.0	140977	9	AC002536 Human Chr
65	16.8	84.0	141449	9	AC129929 Homo sapi

Pred. No. is the number of results predicted by chance to have a

C 66	16.8	84.0	143565	2	AC023195	AC023195 Homo sapi
C 67	16.8	84.0	144461	2	AC145808	Oryza sat
C 68	16.8	84.0	144902	2	AC145808	Oryza sat
C 69	16.8	84.0	145042	2	AC091012	Homo sapi
C 70	16.8	84.0	147505	9	CNS01012	Homo sapi
C 71	16.8	84.0	148083	9	AF111170	Homo sapi
C 72	16.8	84.0	153250	3	AC008332	Drosophila
C 73	16.8	84.0	157566	9	AC022517	Homo sapi
C 74	16.8	84.0	158467	9	AL512556	Homo sapi
C 75	16.8	84.0	162144	9	AL512556	Homo sapi
C 76	16.8	84.0	166380	9	AC087237	Homo sapi
C 77	16.8	84.0	174811	2	AC013259	Homo sapi
C 78	16.8	84.0	181004	9	AC013259	Homo sapi
C 79	16.8	84.0	181683	2	AC021618	Homo sapi
C 80	16.8	84.0	184556	9	AC035021	Homo sapi
C 81	16.8	84.0	190301	3	AC009352	Drosophila
C 82	16.8	84.0	197426	3	CNS07EG9	AL590442 Chromosome
C 83	16.8	84.0	198273	9	AE003638	Homo sapi
C 84	16.8	84.0	254309	3	AE003638	Homo sapi
C 85	16.8	84.0	346362	1	BE640439	Bordetella
C 86	16.8	84.0	348257	1	BE640425	Bordetella
C 87	16.8	84.0	349726	1	BE640421	Bordetella
C 88	16.4	82.0	354	6	AG898401	Sequence
C 89	16.4	82.0	354	6	BD033934	Sequence
C 90	16.4	82.0	674	5	AF095342	Oreophryn
C 91	16.4	82.0	2230	9	BC013116	Homo sapi
C 92	16.4	82.0	2750	8	AK110063	Oryza sat
C 93	16.4	82.0	3145	6	BD275095	47 Human
C 94	16.4	82.0	3287	6	AX301205	Sequence
C 95	16.4	82.0	3345	6	AX405762	Sequence
C 96	16.4	82.0	3349	6	HS4427922	Homo sapi
C 97	16.4	82.0	3370	6	AX833184	Sequence
C 98	16.4	82.0	3370	6	AK094738	Homo sapi
C 99	16.4	82.0	11219	1	AE005981	Caulobact
C 100	16.4	82.0	20000	9	AB006684	Homo sapi
C 101	16.4	82.0	33110	9	AL354671	Human DNA
C 102	16.4	82.0	36284	9	HS409610	Homo sapi
C 103	16.4	82.0	39374	9	AC005256	Homo sapi
C 104	16.4	82.0	39531	2	AF024534	Homo sapi
C 105	16.4	82.0	39782	2	AC138635	Homo sapi
C 106	16.4	82.0	40476	2	AC140538	Homo sapi
C 107	16.4	82.0	4091	9	AC002116	Human DNA
C 108	16.4	82.0	42133	9	AP001060	Homo sapi
C 109	16.4	82.0	43444	2	CH19R31491	Homo sapi
C 110	16.4	82.0	44228	2	AC140533	Homo sapi
C 111	16.4	82.0	44635	2	AC140539	Homo sapi
C 112	16.4	82.0	45185	2	AC011166	Homo sapi
C 113	16.4	82.0	47877	9	HS2060C7	Human DNA
C 114	16.4	82.0	48894	2	AC080102	Homo sapi
C 115	16.4	82.0	6802	2	AC118753	Homo sapi
C 116	16.4	82.0	76142	2	AC139539	Homo sapi
C 117	16.4	82.0	88936	9	AC090830	Homo sapi
C 118	16.4	82.0	88780	2	AC138711	Homo sapi
C 119	16.4	82.0	100976	9	HS1005F21	Human DNA
C 120	16.4	82.0	107467	9	HS958B3	Continuation (6 of
C 121	16.4	82.0	110000	2	AC003656_5	Continuation (6 of
C 122	16.4	82.0	119931	9	AC110285	Homo sapi
C 123	16.4	82.0	129586	9	AC005086	Homo sapi
C 124	16.4	82.0	143192	9	AL954247	Homo sapi
C 125	16.4	82.0	145256	9	AL590226	Human DNA
C 126	16.4	82.0	149567	9	AL590226	Human DNA
C 127	16.4	82.0	158365	9	AC091171	Homo sapi
C 128	16.4	82.0	159064	9	AC138512	Homo sapi
C 129	16.4	82.0	166878	2	AC023288	Homo sapi
C 130	16.4	82.0	167809	2	AC126415	Mus muscu
C 131	16.4	82.0	170959	2	AC022197	Homo sapi
C 132	16.4	82.0	174235	9	AC145919	Fan trogl
C 133	16.4	82.0	175691	2	AC007615	Homo sapi
C 134	16.4	82.0	176502	2	AC134706	Rattus no
C 135	16.4	82.0	177432	9	AC118274	Homo sapi
C 136	16.4	82.0	181714	2	AC090442	Homo sapi
C 137	16.4	82.0	184050	2	AC087819	Homo sapi
C 138	16.4	82.0	184444	2	AC090471	Homo sapi
C 139	16.4	82.0	187824	9	AP001095	Homo sapi
C 140	16.4	82.0	189628	2	AC146318	Gallus ga
C 141	16.4	82.0	191496	2	AC009130	Homo sapi
C 142	16.4	82.0	191541	4	AC144339	Oryctolag
C 143	16.4	82.0	194020	9	AC096679	Pan trogl
C 144	16.4	82.0	195791	2	AC019281	Homo sapi
C 145	16.4	82.0	197675	9	AC096770	Homo sapi
C 146	16.4	82.0	202453	2	AC069094	Homo sapi
C 147	16.4	82.0	203765	2	AC139635	Gallus ga
C 148	16.4	82.0	204433	9	AC094839	Homo sapi
C 149	16.4	82.0	207937	9	AC009083	Homo sapi
C 150	16.4	82.0	216498	2	AC123722	Mus muscu
C 151	16.4	82.0	225204	2	AC099219	Rattus no
C 152	16.4	82.0	226460	2	AC114157	Rattus no
C 153	16.4	82.0	232450	2	AC095706	Rattus no
C 154	16.4	82.0	233479	2	AC126758	Homo sapi
C 155	16.4	82.0	233772	2	AC115403	Rattus no
C 156	16.4	82.0	243676	10	AC115403	Rattus no
C 157	16.4	82.0	256608	2	AC021040	Homo sapi
C 158	16.4	82.0	269607	2	AC145355	Gorilla g
C 159	16.4	82.0	340000	9	AP001753	Homo sapi
C 160	16.4	82.0	340000	9	AP001754	Homo sapi
C 161	16.4	82.0	3417	6	AX481394	Sequence
C 162	16.4	82.0	3845	6	AL6753	tie recepto
C 163	16.4	82.0	3845	6	AR074868	Sequence
C 164	16.4	82.0	3845	6	AX587640	Sequence
C 165	16.4	82.0	3845	6	AX671046	Sequence
C 166	16.4	82.0	3845	9	HSTIEMR	Human tie m
C 167	16.4	82.0	3960	9	BC038239	Homo sapi
C 168	16.4	82.0	6074	9	AF067971	Macaca mu
C 169	16.4	82.0	8670	6	BD129605	Polynucle
C 170	16.4	82.0	8670	8	CRARG7	Chlamydomon
C 171	16.4	82.0	35716	10	AL928570	Homo sapi
C 172	16.4	82.0	91550	2	AL928570	Mouse DNA
C 173	16.4	82.0	100340	9	AC118286	Oryza sat
C 174	16.4	82.0	102286	9	AC087644	Homo sapi
C 175	16.4	82.0	106168	9	AC011450	Homo sapi
C 176	16.4	82.0	129336	9	AC130457	Homo sapi
C 177	16.4	82.0	135417	9	HS4041217	Human DNA
C 178	16.4	82.0	135996	2	AL356007	Homo sapi
C 179	16.4	82.0	140366	9	AC104425	Homo sapi
C 180	16.4	82.0	144988	10	AC101022	Mus muscu
C 181	16.4	82.0	146514	2	AC141983	Rattus no
C 182	16.4	82.0	149480	6	AX329775	Sequence
C 183	16.4	82.0	149480	6	AX329776	Sequence
C 184	16.4	82.0	149480	6	AX336193	Sequence
C 185	16.4	82.0	149480	9	HUU95740	Human Chrom
C 186	16.4	82.0	149480	9	HUU95740	Human Chrom
C 187	16.4	82.0	150129	2	AC023816	Homo sapi
C 188	16.4	82.0	153857	2	AF336380	Mus muscu
C 189	16.4	82.0	168813	9	AC013717	Homo sapi
C 190	16.4	82.0	171016	9	AC104371	Homo sapi
C 191	16.4	82.0	171140	9	AC090685	Homo sapi
C 192	16.4	82.0	172522	2	AC024065	Homo sapi
C 193	16.4	82.0	175611	2	AC020621	Homo sapi
C 194	16.4	82.0	178733	2	AC068786	Homo sapi
C 195	16.4	82.0	181603	9	AC093189	Pan trogl
C 196	16.4	82.0	183067	9	AC040914	Homo sapi
C 197	16.4	82.0	186643	2	AC144995	Rattus no
C 198	16.4	82.0	190183	9	AC100768	Homo sapi
C 199	16.4	82.0	191141	2	AC146230	Pan trogl
C 200	16.4	82.0	191530	2	AC140519	Homo sapi
C 201	16.4	82.0	193766	9	AC093420	Homo sapi
C 202	16.4	82.0	194981	9	AC145997	Pan trogl
C 203	16.4	82.0	210688	2	AC139257	Homo sapi
C 204	16.4	82.0	212387	9	AC140504	Homo sapi
C 205	16.4	82.0	213447	2	AC13473	Mus muscu
C 206	16.4	82.0	220641	2	AC147123	Pan trogl
C 207	16.4	82.0	221062	2	AC098570	Mus muscu
C 208	16.4	82.0	226841	2	HSAC002043	Homo sapi
C 209	16.4	82.0	240453	2	AC119387	Rattus no
C 210	16.4	82.0	240453	2	AC119387	Rattus no
C 211	16.4	82.0	246240	2	AC134895	Rattus no

212	16	80.0	250608	2	AC110700	Rattus no
213	16	80.0	346259	1	BX540435	Bordetell
214	15.8	79.0	454	6	AX577678	Sequence
215	15.8	79.0	477	6	AR427389	Sequence
216	15.8	79.0	477	6	BD122942	EST and e
217	15.8	79.0	504	6	AX577682	Sequence
218	15.8	79.0	523	6	AX873301	Sequence
219	15.8	79.0	523	6	BD153963	Primer fo
220	15.8	79.0	578	9	HSXEBV1	
221	15.8	79.0	598	8	AX064351	H.sapiens X
222	15.8	79.0	636	9	HUMU73B11	
223	15.8	79.0	786	6	AX474216	Sequence
224	15.8	79.0	786	6	AX474218	Sequence
225	15.8	79.0	786	6	AX474220	Sequence
226	15.8	79.0	786	6	AX474222	Sequence
227	15.8	79.0	786	6	AX474234	Sequence
228	15.8	79.0	786	6	AX474237	Sequence
229	15.8	79.0	786	6	BD185471	Peptide m
230	15.8	79.0	786	6	BD185472	Peptide m
231	15.8	79.0	786	6	BD185473	Peptide m
232	15.8	79.0	786	6	BD185474	Peptide m
233	15.8	79.0	786	6	BD185480	Peptide m
234	15.8	79.0	786	6	BD185482	Peptide m
235	15.8	79.0	836	3	CGU40207	Sequence
236	15.8	79.0	947	6	AX367930	Sequence
237	15.8	79.0	947	6	AX069385	Sequence
238	15.8	79.0	947	6	AX335617	Sequence
239	15.8	79.0	947	6	AX410772	Sequence
240	15.8	79.0	947	9	HSU73514	Human short
241	15.8	79.0	960	6	AX053564	Sequence
242	15.8	79.0	961	9	BC000372	Homo sapi
243	15.8	79.0	965	9	BC008708	Homo sapi
244	15.8	79.0	968	9	AF069134	Homo sapi
245	15.8	79.0	973	9	AF035555	Homo sapi
246	15.8	79.0	990	9	HSU96132	Sequence
247	15.8	79.0	1005	6	AX053562	Sequence
248	15.8	79.0	1054	6	BD082375	Sequence
249	15.8	79.0	1193	1	AX346077	Trepnema
250	15.8	79.0	1568	6	AX577680	Sequence
251	15.8	79.0	1581	9	AK025867	Homo sapi
252	15.8	79.0	1602	6	AX577684	Sequence
253	15.8	79.0	1638	10	BC028637	Mus muscu
254	15.8	79.0	1655	3	CGY220170	Sequence
255	15.8	79.0	1692	9	BC028014	Homo sapi
256	15.8	79.0	1703	3	VS0270180	Vannucula
257	15.8	79.0	1704	3	CS0270168	Calvicia
258	15.8	79.0	1749	9	BC034321	Homo sapi
259	15.8	79.0	1877	9	BC004526	Homo sapi
260	15.8	79.0	2064	6	AX834571	Sequence
261	15.8	79.0	2064	9	AK097109	Sequence
262	15.8	79.0	2431	1	MSGBCG	Homo sapi
263	15.8	79.0	2531	9	HSNDSADH	M1705 M.bovis BCG
264	15.8	79.0	2597	9	HSAL17866	Y11192 H.sapiens m
265	15.8	79.0	2689	10	AK122329	Homo sapien
266	15.8	79.0	2731	6	AX746759	Mus muscu
267	15.8	79.0	2731	9	AK091062	Sequence
268	15.8	79.0	2747	6	AK877398	Homo sapi
269	15.8	79.0	2747	6	BD156637	Sequence
270	15.8	79.0	2747	9	AK001129	Primer fo
271	15.8	79.0	2874	6	AX880200	Homo sapi
272	15.8	79.0	2874	6	BD158251	Sequence
273	15.8	79.0	2874	9	AK027636	Primer fo
274	15.8	79.0	3396	8	AK109980	Homo sapi
275	15.8	79.0	3926	9	HSN801327	Oryza sat
276	15.8	79.0	4079	9	AF037438	Homo sapi
277	15.8	79.0	4380	1	MSGTCWPA	AF037438 Homo sapi
278	15.8	79.0	4380	6	I08847	M15467 M.tuberculo
279	15.8	79.0	4380	6	I08856	Sequence 5
280	15.8	79.0	4380	6	AR262811	Sequence 10
281	15.8	79.0	4976	9	AB009577	Sequence
282	15.8	79.0	4997	9	HSN805783	Homo sapi
283	15.8	79.0	5054	9	AB046853	Homo sapi
284	15.8	79.0	5103	9	HSAA427355	Homo sapi

RESULT 1

AE007080

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

1. 17131

2. (bases 1 to 17131)

3. (bases 1 to 17131)

4. (bases 1 to 17131)

5. (bases 1 to 17131)

6. (bases 1 to 17131)

7. (bases 1 to 17131)

8. (bases 1 to 17131)

9. (bases 1 to 17131)

10. (bases 1 to 17131)

11. (bases 1 to 17131)

12. (bases 1 to 17131)

13. (bases 1 to 17131)

14. (bases 1 to 17131)

15. (bases 1 to 17131)

16. (bases 1 to 17131)

17. (bases 1 to 17131)

18. (bases 1 to 17131)

19. (bases 1 to 17131)

20. (bases 1 to 17131)

21. (bases 1 to 17131)

22. (bases 1 to 17131)

23. (bases 1 to 17131)

24. (bases 1 to 17131)

25. (bases 1 to 17131)

26. (bases 1 to 17131)

27. (bases 1 to 17131)

28. (bases 1 to 17131)

29. (bases 1 to 17131)

30. (bases 1 to 17131)

31. (bases 1 to 17131)

32. (bases 1 to 17131)

33. (bases 1 to 17131)

34. (bases 1 to 17131)

35. (bases 1 to 17131)

36. (bases 1 to 17131)

37. (bases 1 to 17131)

38. (bases 1 to 17131)

39. (bases 1 to 17131)

40. (bases 1 to 17131)

41. (bases 1 to 17131)

42. (bases 1 to 17131)

43. (bases 1 to 17131)

44. (bases 1 to 17131)

45. (bases 1 to 17131)

46. (bases 1 to 17131)

47. (bases 1 to 17131)

48. (bases 1 to 17131)

49. (bases 1 to 17131)

50. (bases 1 to 17131)

ALIGNMENTS

17131 bp DNA linear BCT 27-APR-2001

AE007080 Mycobacterium tuberculosis CDC1551, section 166 of 280 of the

complete genome.

AE007080 AE000516

AE007080.1 GI:13982094

Mycobacterium tuberculosis CDC1551

Mycobacterium tuberculosis CDC1551

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium

tuberculosis complex

1 (bases 1 to 17131)

Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,

Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,

Kolony,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,

Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,

Gill,J., Mikula,A. and Bishai,W.

Whole genome comparison of Mycobacterium tuberculosis clinical and

laboratory strains

Unpublished

2 (bases 1 to 17131)

Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,

Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,

Kolony,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,

Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,

Gill,J., Mikula,A. and Bishai,W.

Direct Submission

Submitted (25-APR-2001) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers

1. 17131

/organism="Mycobacterium tuberculosis CDC1551"

/mol\_type="genomic DNA"

/strain="CDC1551"

/db\_xref="taxon:83331"

/notes="clinical strain"

104. 1357

/gene="MT2390"

104. 1357

/gene="MT2390"

/note="similar to SP:P71894, PID:1524272 GB:AL123456;

identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="PE family protein"

/protein\_id="AAK4682.1"

/db\_xref="GI:1382095"

/translation="MTANANLARKORHFTTINVRSSLNNRSESRGTSMQELSVIPE

QVESAAQDLAIGRSALSAAGPTTAVSAADEVSTAASIFGAYGRQCQLISA

QASAFHDEFVNLTKGATYRNTAFANAQSNVLNAVNPASRLIGHPSAAESVQNSAP

```

TLGGHSTVITAGLAQAQRAVATVEQAAAVAPLPSAGAGLAQVWNGVVTAGGSA
KLATALQSAWPLAKSGGEFIVAGSALTGVALLQPAVVQVQAGGTFLTAGTSAATG
LGLTLGAGFSGQVGNLALASGTAATGLGLSGAGVQLFSPAFLAVPTLGGVSL
AIAYVQLVQGVQHLVLPVNVVAGTAALQTAGAQAQVAGVNHMTLAAQLGAPGIAVLQT
AGGHPAQGGIGHLTATAGNAATVVLIS"
complement (1392..2939)
/gene="MT2339"
complement (1392..2939)
/gene="MT2339"
/note="similar to PID:1813569; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="nitrite extrusion protein"
/protein_id="AAK46683.1"
/db_xref="GI:13882096"
/translation="MEQHTLLQREBSPSPAASURRLGGSRHITHDPEDLGAWEAR
NGIARNLNLSVTVTHLGSVMTLPVLELLMQDVYGFSTSKFLLGTLATLFGAF
LMPYALASAI FGGNWFATSAI VLLI PAIGTTLVTHPGLPLFYLCAALTGLGG
NFASMSNANAFY PHELKGSALGAGVNLGPALQVGLLAIATVGERPYLVGAL
YVULVAIVGVSLFMNVVEQHVQVNRLRPIVSALSTRTWLALLYLGLTFGSGFIG
FVFGQVLQTMFLACGQSPARATLHVELAFVGPLLAARVILYGGRLADNGSRLT
LIVFVMTLUALGLISASTLEGRHVQGRGATWGVFVCFVALFVLSGLNGSVYKMI
PTIFACSRSLDSLEARDWSRIISGVVGFVAAPGALGGVGNLNAKRESYLSGSG
TDAFWIFMVCYAAAALVTKVYDRRTVTDGMGLQALVRQPASTPAELIGRTQSDRF
SGCSISA"
complement (3173..3700)
/gene="MT2339"
complement (3173..3700)
/gene="MT2339"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46684.1"
/db_xref="GI:13882097"
/translation="MRRQSAVPIALALLALLALLIVLGLSGCAWKPTRRSPN
TKDSGPDADVRAQAAVPIVWPGSKWEITRGHTRNCRHLWQIIPITIASSTPQ
QLIFFRNPPLGSPTRNRPYITLVPAGDDTVTVQWQIGSDQCECPTGIGTVRPHI
GSDGKLEALGSIPIHQ"
3574..4161
/gene="MT2339"
3574..4161
/gene="MT2339"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46685.1"
/db_xref="GI:13882098"
/translation="WTVVLSAASRRNRPPAQSPVPTVPTVPESGRRYADARH
EISENWEFFLFINAGVDPKARVPVFLPQIGRLTPDPAAGEAIGAADIPMAARWI
GRSPCLIGQNTMGDMGLGCLGAGQRCVDRLVNGASRSTCSRLPVIASYDERLSV
LKEVRLHSISIFIKRPGCEVLTVTGYNFRGP"
4596..6232
/gene="MT2394"
4586..6232
/gene="MT2394"
/note="similar to GB:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="malate oxidoreductase"
/protein_id="AAK46686.1"
/db_xref="GI:13882099"
/translation="MSDARVPRI PAALSAPSINRNGVGFTHAQRRLGLTGRPLSAVLT
LDQAEVWQQLATLGRNLLLEQLHYHEVLYFKVLAHLPELMPVVTPTVGE
AIQRFDEYRQGLFLSIDEPDEIEEAFNTLGLGPEVDVLDICTDAEALIGDWGV
GGIQAIVGKALYTAGGVDPDRCLAVSDVGTNDNEQLLADPFYLGNRHARRGREYD
EFVSRVLEAQLFPRAILHFEDEFGANARKILDTYDTCVFNDDMOGTQAVLAAV
YSLKVTGTGIPLDQTVFPGACTAGNADQIRDAWADGATLEOASQIWPIDRPLG
LFDMDMDLDFQVPYAKNRHLQGVAGVDRVGLSDAIKIASPILIGCSTVYGAFTKEV
VEAMTASCKHPMIFPUSNPSTSRMEALPADVLAWSNGRALLATGSPVAPVEFDETVVI

```

```

GOANNVLAPPGIGLIVAGARLITRMLHAAAKAIAHOANPTNPGDSILLPDVQNLRA
ISITVAEYVRAVQGVASRTHDDVYRQAI VDTMWLPAYD"
complement (6186..7799)
/gene="MT2395"
complement (6186..7799)
/gene="MT2395"
/note="similar to GP:5881853; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="drug transporter"
/protein_id="AAK46687.1"
/db_xref="GI:13882100"
/translation="NNRTQLTLIATGLFMIFLDALIVNVALPDIQSFVAGBEDGL
QWVVASVLSGMVTFINSAATLADLYGRRRWYLGVSFLTGLSGIACGLAPSIAVLITAR
GAOGLGAAAVSTSLAVSAARPEAKERAKAIGITWAIASIGTTTGTGLGLVDQWG
WRSIFYNLPKMGALVFLCYVVEESCNERARFPDLSGOLLFIYAVGALVAVIEGPQ
IGWTSQITVLMUATAVGCALFVWLERRSNPMDLTDRDTSYALAIAITICTVPEAV
YGMLLTTOFLQNVRYTPTVGLMILPFSAAVIVSPVLVGLVGRIGARVPILAGLC
MLMLGLMIFSEHRSALVGLGCGSVALCLPTITTVATAVPAERAGASGIM
SAQRAICSTIGFAVGLSVAAMLSATLEPHEAEVDPVQVHVLABIIIDSANPRAHV
GGIVPRRHIEHRDPAIABEDFTIGIRVALLVATA"LAUVFLAGMWFPRDHTAGSD
LSRLPTAMTVECAVSHMPGATWCRWPA"
7826..7996
/gene="MT2396"
7826..7996
/gene="MT2396"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46688.1"
/db_xref="GI:13882101"
/translation="MAICDQTLGCASDGVSRFGRFEPFNGRRLNRPNQWGRKFTAGVS
QLGEPPLAHL"
8274..9206
/gene="MT2397"
8274..9206
/gene="MT2397"
/note="similar to PID:2342631; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="cysteine synthase"
/protein_id="AAK46689.1"
/db_xref="GI:13882102"
/translation="NSIADITOLIGRTPLVLRVRRVTDGAVADIVAKLEFFNPANSVK
DRIGVAMLQAEQGLIKPDITILEPTSGNTGIALVCAAGYRCVLTMPETMSLER
RMLRAYGAELITPGADGMSGAIAKAEELAKTDQRYFVFPQFPANPAIHRVTAAE
EVRWTDKVDIVVAGVGTGTTVAQVIKERKSFARFVAFVPAASPVLSGGQKGFH
PIQIGAGFVPPVLDODLVDEIITVGNEDALNVARRLAREBGLLVIGISSAATAVAALQ
VARRPENAGLIVVLPDFGERVLSPLFADVAD"
9210..9899
/gene="MT2398"
9210..9899
/gene="MT2398"
/note="similar to GP:2342632; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="serine acetyltransferase CysB, putative"
/protein_id="AAK46690.1"
/db_xref="GI:13882103"
/translation="MLTAMEGDIRAARERDPAAPTALEVIFCYGVHVMGHRHLAHL
WQGARILARAARAEFTRLTGVDIHPGAVIGARVFDHATGVVIGTAEVGDVITYH
GVTLGSGMVGKGRHPTVGDRIIGAGAKVIGPIKIGEDSRICANAVVVKVPVSAVV
VGVPQVIGSQSPSGPFPDWRLPDLVGASLDSLLTRVARELDALGCGPQAAGVIRPPE
AGIWHGDEFSI"
10315..11283
/gene="MT2399"
10315..11283
/gene="MT2399"
/note="identified by Glimmer2; putative"

```



```
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AK46691.1"

Query Match          95.0%; Score 19; DB 1; Length 17131;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2  GAGGGCCAGCCAGCCACAA 20
      |||||
DB      3615 GAGGGCCAGCCAGCCACAA 3633

RESULT 2
BX248342
LOCUS      BX248342
DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment
9/14.
ACCESSION BX248342 BX248333
VERSION    BX248342.1 GI:31619031
KEYWORDS   complete genome.
SOURCE     Mycobacterium bovis subsp. bovis AF2122/97
ORGANISM   Mycobacterium bovis subsp. bovis AF2122/97
            Bacteria: Actinobacteria: Actinobacteridae; Actinomycetales:
            Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
            tuberculosis complex.
REFERENCE  1
AUTHORS    Garnier,T., Eiglmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
            Pryor,M., Duthey,S., Grondin,S., Lacroix,C., Monsemp,C., Simon,S.,
            Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,
            Wheeler,P.R., Parkhill,J., Barrell,B.G., Cole,S.T., Gordon,S.V. and
            Hewinson,G.
            The complete genome sequence of Mycobacterium bovis
            Online Publication
            PNAS 10.1073/pnas.1130426100 ( Microbiology )
            2 (bases 1 to 306550)
            Garnier,T.
            Direct Submission
            Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
            Bacterienne Institut Pasteur,28,rue du Dr Roux 75724 PARIS cedex
            15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
            Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
            Laboratories Agency Weybridge, Woodham Lane, New Haw Addlestone,
            Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK. F44 Annotation, Genecopie, Institut
            Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
            Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
            rue du Docteur Roux, 75724 Paris Cedex 15, France
            Location/Qualifiers
            1. 306550
            /organism="Mycobacterium bovis subsp. bovis AF2122/97"
            /mol_type="genomic DNA"
            /strain="AF2122/97"
            /db_xref="taxon:233413"
            201..1286
            /gene="adhE2"
            /locus_tag="Mb2283"
            201..1286
            /gene="adhE2"
            /locus_tag="Mb2283"
            /EC_number="1.3.1.-"
            /note="Mb2283, adhE2, len: 361 aa. Equivalent to Rv2259,
            len: 361 aa, from Mycobacterium tuberculosis strain H37Rv,
            (100.0% identity in 361 aa overlap). Probable adhE2,
            zinc-containing alcohol dehydrogenase, similar to several,
            especially mycothiol-dependent formaldehyde dehydrogenase
            from Amycolatopsis methanolica P80094 (360 aa). Contains
            P500059 Zinc-containing alcohol dehydrogenases signature.
            FASTA scores: >sp|P80094|FADH AMYME
            NAD/MYCOTHOL-DEPENDENT FORMALDEHYDE DEHYDROGENASE
            (MD-FALDH) Length = 360, Expect = e-156, Identities =
            268/358 (74%). Also similar to Rv0162c, (MTC128.02c, 35.0%
```

```
identity in 371 aa overlap)."
/codon_start=1
/transl_table=11
/product="Probable zinc-dependent alcohol dehydrogenase
adhE2"
/protein_id="CAD97144.1"
/db_xref="GI:31619032"
/translations="MSQTVRGVIARQKGFPELVNIVVPDPGGEAVVDVTACGVRGT
DLVREGGINDEYFPLLGHEAAGIIEAVGVGVTAVERGDFVILNMRVACQCRACKRG
RPRVCFDTNABEOKMTLTDGTELTAAIGAFADKTLVHSGQCTKVPADPAVAGLL
GCGVMAGLNAINTGGVTRDDTVAVIGCGGVGDAIAGALVGAKRIIAVDTDDTKLD
WARTFGAHTVNAKREVDVVOAIGLIDFGADVVIDAVGRPEFYQQAIFYARDIAGTIV
LVGVPTDMRLDMLPVDFFSHGGALSGWSYGDCLPESDFTLIDLVLQGLRPLQRFVS
ERICLEDVEEAFPHKMGKGLRSVVM"
1286..1921
/locus_tag="Mb2284"
1286..1921
/locus_tag="Mb2284"
/note="Mb2284, -, len: 211 aa. Equivalent to Rv2260, len:
211 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 211 aa overlap). Conserved
hypothetical protein, similar to hypothetical proteins
Rv0634c, Rv1637c, Rv3677c, Rv2581c from Mycobacterium
tuberculosis and to various hydrolases. FASTA scores:
spr|006154|006154 HYPOTHETICAL 21.3 KD PROTEIN (200 aa)
opt: 355, E(): 4e-15; (37.4% identity in 198 aa
overlap)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD97145.1"
/db_xref="GI:31619033"
/translations="MAAERVITHGFELDGGSWENVNINLWGDSEVVVFAAHHHA
APIIDAVGGRKVVAVICTGHNDHVTVAPELGTALDAPVLHFGDAPVLMHTPKDSF
RAYSDDGAVRVGTELRALHTPGHSPGVCWYAPELGPTGTGTFSGDTGFAGGPGATG
RSYSDPFTILRSISGRIGALPGDVTVHTGDSGTTCIDETVHYEEVWARGH"
complement(1998..3506)
/locus_tag="Mb2285c"
complement(1998..3506)
/locus_tag="Mb2285c"
/note="Mb2285c, -, len: 502 aa. Equivalent to Rv2262c and
Rv2261c, len: 360 aa and 140 aa, from Mycobacterium
tuberculosis strain H37Rv, (94.7% identity in 357 aa
overlap and 100.0% identity in 140 aa overlap). Conserved
hypothetical protein, with function unknown but some
similarity to N-terminal 70% of
P23930|P77703|LMT_ECOLI|CUTE|E0657 APOLIPOPROTEIN
N-ACYLTRANSFERASE [EC 2.3.1.-] from Escherichia coli
strain K12 (512 aa), FASTA scores: opt: 239, E(): 1.6e-07,
(30.4% identity in 359 aa overlap). Note that neighboring
ORF shows similarity to N-terminal part of PCC6803
apolipoprotein N-acyltransferase from Synechocystis sp.,
suggesting possibility of frameshift. Sequence of clones
from two sources has been checked but no error found.
Appear to be two extra bases at position 1876970 compared
to CDC1551 strain. Conserved hypothetical protein, with
function unknown but some similarity to C-terminal end of
PCC6803 apolipoprotein N-acyltransferase from
Synechocystis sp. Note that next ORF shows similarity to
N-terminal part of P74055 APOLIPOPROTEIN N-ACYLTRANSFERASE
from Escherichia coli (519 aa), FASTA scores: opt: 142,
E(): 0.007, (29.9% identity in 117 aa overlap), suggesting
possible frameshift. Sequence of clones from two sources
has been checked but no error found.
REMARK-M.bovis-M.tuberculosis: In Mycobacterium
tuberculosis strain H37Rv, Rv2262c and Rv2261c exist as 2
genes. In Mycobacterium bovis, a 2 bp deletion (ct-*)
results in a single product which is more similar to
Rv2262c."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD97146.1"
```

```

/db_xref="GI:31619034"
/translation="MALRAGARQPVIGCAALVFGALPFAFPAPSWWNLAWFLGVLV
LILVVAARLAVGWAGTGVFLATQVWLTSGPMLVLLAAGLGVLLPAGWL
AHRLLSVPTTCVGAALVVPASVAAEAVFSWOSLGPWALLGASOWSOPVTLASA
SIGGWLTSFLVAVNTALIASVLVCRATGGRVLGCVIGCAGLGPASVILGSPVGS
PTRVVALVQAGDIAAARAAAGBEFTAAVADQRPDLVVGESSVGGQDLTRHPDVLAR
LAELSQRVADLLVNVDPADPGGYKGAVLVGAHEAVGKRLKLPFGYVPLRPL
FGWITSTFAAARQKRGAPVTVLAVNSLHAPLISYEMTFSDLTRHAARLGAALLVY
QGSTSTFOGSAQPOAQAQPAVRAVEAGIPAVHASLSDSSAFDTRGRRLAWCSAEFN
GALVNVPLASNVTLVRLGDWVPVPTAFVVMGAGFAVFLRRSLARVSDCAK"
3595..4548
/locus_tag="Mb2286"
3595..4548
/locus_tag="Mb2286"
/notes="Mb2286", len: 317 aa. Equivalent to Rv2263, len:
317 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 317 aa overlap). Possible
oxidoreductase (EC 1.1.1.1), similar to several
oxidoreductases. Similarity suggests alternative GTG start
at 10154 but then no rbs. FASTA scores: sptr|Q544
05|Q54405 PROBABLY AN NADP-DEPENDENT OXIDOREDUCTASE (297
aa) opt: 487, E(): 1.1e-23; (36.1% identity in 299 aa
overlap). Also similar to M. tuberculosis Rv0068, and
Rv0439c.
/codon_start=1
/transl_table=11
/product="Possible oxidoreductase"
/protein_id="CAD97147.1"
/db_xref="GI:31619035"
/translation="NAKDIATVATVPLDSGLKLAITGANGSLGFLARRLSAAGAVIMA
INRAGEAAVEITVAPDAKLTALDLSLVAALGOLMADGSPIDILLINAG
VMPPRVITAGFELQFQSGHLGHFALTPHLLPRAQARVVSLSLRAGEIH
FDLPERSVAPMTYAGQKLAFLMFARELDKRSAGVINSNAHGLKTNLQIA
GSHGSDPALMEYKTSWRFAPLQWIEGILPALLYAATPOADGGAFYGRGRI
EVAGGVRKAPAAARNADSKRLMEVSEQLTGVSYPKSR"
complement(4526..6304)
/locus_tag="Mb2287c"
complement(4526..6304)
/locus_tag="Mb2287c"
/notes="Mb2287c", len: 592 aa. Equivalent to Rv2264c,
len: 592 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 592 aa overlap). Conserved
hypothetical Pro-rich protein, similar to hypothetical
proteins Rv0312 (MTCY63.17, 620 aa and Rv0350) that has
highly Pro-, Thr-rich C-terminus. Contains PS00343
Gram-positive cocci surface proteins 'anchoring'
hexapeptide. FASTA scores: Z96800|MTCY63.17 Mycobacterium
tuberculosis cosmid (620 aa) opt: 1075, E(): 8.8e-24;
(38.9% identity in 627 aa overlap).
/codon_start=1
/transl_table=11
/product="conserved hypothetical proline rich protein"
/protein_id="CAD97148.1"
/db_xref="GI:31619036"
/translation="MATGARPALGSGVNTNLAAVAADHSITRKPVLTILYRQRPPEVG
LPSENRLDEPGVITDFVRVGDGSGIVAADGVSYSALVADALLAYATATGTRA
VPGSVTVTPAHMGPAVAALDALSARSEWGHSTQAQPLSLPDAALYAIRAD
PGIPARGIVACDVGSGGTGTLVDAADYVRVAATVRHQAFSGDLIDQSLISVWSE
LPGTGADPAGTSAIGSLKTRIECKKARLSSTVTTLTALGGLDILTRNELEDT
IRDSLSVGALEQLTLAGSIRTAELVALVSVGGGANIPAVTTTSGSPCVVPTPR
PQUTAPFGALMAARPGDTSATLVATVSATATAPADAPASVLQALWSEADSDSH
IGPAPYAAARSLSDHDAEPKSPPIPIWYRLPAVITGTTIVALLVWSEADSVIG
LSTGDQPTAFQTPRGVTTTAAPEPSAPDTPPEPAPVPAPATGGPAPQOP
LPPTPTTTTQPAVTTTTPATPTPASAPPATTPPTTQPPATTSPPSPPIPIPP
IPBIPOLPGIPQVPGIQFSAISGS"
6654..7883
/locus_tag="Mb2288"
6654..7883
/locus_tag="Mb2288"

```

Query Match 95.0%; Score 19; DB 1; Length 306550;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches	19;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	2	GAGGCGCCAGCCCCACAA	20						
Db	70732	GAGGCGCCAGCCCCACAA	70750						
RESULT 3									
EX842579		standard; circular genomic DNA; PRO; 348247 BP.							
ID	EX842579								
XX	AC	EX842579; AL021246; AL021925; AL021957; AL123456; Z70283; Z70692; Z77163;							
XX	AC	Z79702; Z81368; Z81371; Z81451; Z83860; Z95388;							
XX	SV	EX842579.1							
XX	XX	21-NOV-2003 (Rel. 77, Created)							
DT	DT	21-NOV-2003 (Rel. 77, Last updated, Version 1)							
XX	XX	Mycobacterium tuberculosis H37Rv complete genome; segment 8/13							
DE	XX	complete genome.							
XX	XX	Mycobacterium tuberculosis H37Rv							
OS	XX	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;							
OC	XX	Corynebacterineae; Mycobacteriaceae; Mycobacterium;							
OC	XX	Mycobacterium tuberculosis complex; Mycobacterium tuberculosis.							
XX	XX	[1]							
XX	XX	MEDLINE; 98295987.							
XX	XX	PUBMED; 9634230.							
XX	XX	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,							
XX	XX	Gordon S.V., Eiglmeier K., Gas S., Barry III C.E., Tekala F., Badcock K.,							
XX	XX	Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K.,							
XX	XX	Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K.,							
XX	XX	Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J.,							
XX	XX	Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S.,							
XX	XX	Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;							
XX	XX	"Deciphering the biology of Mycobacterium tuberculosis from the complete							
XX	XX	genome sequence";							
XX	XX	Nature 393:537-544 (1998).							
XX	XX	[2]							
XX	XX	PUBMED; 12368430.							
XX	XX	Canus J.C., Pryor M.J., Medigue C., Cole S.T.;							
XX	XX	"Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv";							
XX	XX	Microbiology 148:2967-2973(2002).							
XX	XX	[3]							
XX	XX	Submitted (11-JUN-1998) to the EMBL/GenBank/DBJ databases.							
XX	XX	Submitted on behalf of the Mycobacterium tuberculosis sequencing and							
XX	XX	mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton,							
XX	XX	Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut							
XX	XX	Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail:							
XX	XX	parkhill@sanger.ac.uk							
XX	XX	Notes:							
XX	XX	Details of M. tuberculosis sequencing at the Sanger Centre							
XX	XX	are available on the World Wide Web.							
XX	XX	(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/)							
XX	XX	Key							
XX	XX	Location/Qualifiers							
XX	XX	1..348247							
XX	XX	/db_xref="taxon:83332"							
XX	XX	/mol_type="genomic DNA"							
XX	XX	/organism="Mycobacterium tuberculosis H37Rv"							
XX	XX	/strain="H37Rv"							
XX	XX	/complement(209..2248)							
XX	XX	/note="Rv2163c, (MTCY270.05), len: 679 aa. Probable pbpB,							
XX	XX	CDS							

```
FT penicillin-binding membrane protein, similar to many
FT bacterial PBP2 proteins e.g.
FT P11882|PBP2_NEIME|PENA|NMA2072|NMB0413|penicillin-binding
FT protein 2 (pbb-2) from Neisseria meningitidis (serogroups A
FT and B) (581 aa), FASTA scores: opt: 665, E(): 1.6e-31,
FT (33.2% identity in 591 aa overlap); etc. Also similar to
FT Rv0046c and Rv2864c from Mycobacterium tuberculosis
FT (2.8e-10). Contains P80017 possible ATP/GTP-binding site
FT motif A (P-loop) near C-terminus. FASTA best: PBP2_NEIME
FT P11882|penicillin-binding protein 2 (pbb-2). (591 aa) opt:
FT 665, E(): 1.6e-31; (33.2% identity in 591 aa overlap)"
FT /transl_table=11
FT /function="Involved in peptidoglycan biosynthesis."
FT /gene="pbbB"
FT /locus_tag="Rv2163c"
FT /product="Probable penicillin-binding membrane protein
FT pbbB"
FT /protein_id="CAB08664.1"
FT /translation="MSRAAPRASQSTRPARGLRPPGAGEVGORCPGKTKAKOAA
FT QBATSRATRSADVAPAGSIPARTKTRQVVDVGTGASFEVFRHRTGNAILVLMVAAT
FT FTQLQVSHAGRAQAQKLVTDVQPAAGSIVDRNRDLAFTIEARLTFQPKRIR
FT ROLESARKTSAPQQRDLRIAQEAANKNKPDAAVLKGLSDSTFFYLARADTG
FT AVASAIKAPYVGAERCDLRQYPGSLAANVVGIDWDHGLGLEDLSDLAVLAGTDG
FT SVYDRGSDGVIPGSYRNHRKAVHGSTVLTLDNDIOFYVQQVQQAKNLSGAHNSVA
FT VVLDAKTGEVLAMNDNTPDPSODIGROGDKLGNPVSPPPEGSVKNIVAASAVTEH
FT GLSSDEVLQVPGS|OMGSGVTVHDAWEGHVMPTTGVGKSSNVGTLMLSORVGERY
FT YDMLKFLGQRTGVLPGESAGLVPPIDQWSGSTFANLPICQGLSMTLMTQMYQAI
FT ANDGVRPPIKIAVADPGSTREPRDIDRVSQAQIAVRQMLRAVQDDPMGYQQ
FT GTGPGVAGVYQMAQGTAGQINPCGCGYDFDVYMTIFAGIATADNPRYVIGMLNDP
FT AANSQGAHSAAPLFIHNIAGLMQRENVPLSPDGPPLVLAQT"
FT complement (470). .493)
FT /note="P80017 ATP/GTP-binding site motif A (P-loop)"
FT /note="complement (2245). .3399)
FT /note="Rv2164c, (MTCV270.04), len: 384 aa. Probable pro-
FT rich conserved membrane protein, equivalent to
FT MUC907|AL022602|putative conserved membrane protein from
FT Mycobacterium leprae (377 aa) (AL022602), FASTA scores:
FT opt: 1495, E(): 1.7e-56, (62.217% identity in 397 aa
FT overlap)."
FT /transl_table=11
FT /function="UNKNOWN"
FT /locus_tag="Rv2164c"
FT /product="PROBABLE CONSERVED PROLINE RICH MEMBRANE PROTEIN"
FT /protein_id="CAB08663.1"
FT /translation="MRKREAPKSRSSRRRRADSPAAATRTTNSAPSRIRSRACK
FT TSAPGRQVRSPGPTSPMLSPDPAPAKNTSOAKARAKAKAPKIVPTPMERL
FT AARLSIDRLPLANKYFVVLVIGSLGVGLTLWLSTDAAERSYLSNARETNML
FT QOKKALDRVREASALABAAARQMIETRDFAHLVQDPDGNWVVVGTPEADGVF
FT PPLNKKUPEDPPPPKAAVLEVFVRVTPGDDPAPPAKSGBEVLVTRTPDGTATLGG
FT ATHLPTQAPQLPGVPPIPGAPGVPEPLGAPSPAPENPVLQVGAAPAGLPGPA
FT VPAATPGLSGSGQPMVAPPVAPVANGEGFVPTAPVPTAPAPR"
FT complement (3396). .4586)
FT /note="Rv2165c, (MTCV270.03), len: 396 aa. Conserved
FT hypothetical protein, shows strong similarity to several
FT hypothetical bacterial proteins but has extra 80 aa
FT residues at N-terminus FASTA best: YLXA_BACSU 007876
FT hypothetical 35.3 kDa protein in ftsl (311 aa) opt: 781,
FT E(): 0; (45.6% identity in 296 aa overlap), BELONGS TO THE
FT YABC (E.COLI), YLXA (B.SUBTILIS) FAMILY"
FT /transl_table=11
FT /function="UNKNOWN"
FT /locus_tag="Rv2165c"
FT /product="CONSERVED HYPOTHETICAL PROTEIN"
FT /protein_id="CAB08662.1"
FT /translation="MQTRAPWSLPTALAYFPNARFVSSDRDLGAGAPGIAASRSTAC
FT QTWGGITVADPGSGTGHVPLVLAQSRFELLTALTRYPDGSOAVLLDAGAGHA
FT ERFGLGLPGLRIGLDPTALDVARSRLVRFADRLTAVHTRPCLGALAEISGVAAGV
FT SVDGILFDLGVSMOLDRAEGRGVAAPLDMDMWDPTPLTAADLVNVTYDEALAIL
FT RYRGERRARRIAGGIVRRRAKTPPTSTREIVALLYQAIAPAPRVGGHPAKRTFQALR
FT IAVNDELSRTAVPAADALAGIRIAVLAYQSLDRIKVRVFAEVAATPAGLPVE
FT LPGHPRFRSTHGAERASVAEIERPSTVRRLALQVRHEHRAQSQOWATEKGS"
FT complement (4588). .5019)
FT /note="Rv2166c, (MTCV270.02), len: 143 aa. Conserved
FT CDS
```

```
FT hypothetical protein; shows strong similarity to several
FT hypothetical bacterial proteins such as YLLB_BACSU P55343.
FT Is equivalent to Mycobacterium leprae hypothetical protein
FT ML0905 (143 aa, 92% identity) MLCB268.11c
FT >sp|O09563|YL66 MYCIE HYPOTHETICAL 16.1 KDA PROTEIN ML0905
FT >gi|3080482|emb|CAA18677.1|(AL022602)
FT >gi|13092975|emb|CAC1286.1|(AL0583920). FASTA scores:
FT ML0905|ML0905 conserved hypothetical protein (143 aa) opt:
FT 873, E(): 3.1e-52; 92.254% identity in 142 aa overlap;
FT YLLB_BACSU P55343 hypothetical 16.6 kDa protein (143 aa)
FT opt: 340, E(): 3.6e-17; (35.0% identity in 143 aa overlap).
FT BELONGS TO THE YABB (E.COLI), YLLB (B.SUBTILIS), MG221 (
FT X.GENTILIUM) FAMILY"
FT /transl_table=11
FT /function="UNKNOWN"
FT /locus_tag="Rv2166c"
FT /product="CONSERVED HYPOTHETICAL PROTEIN"
FT /protein_id="CAB08661.1"
FT /translation="MELGTYTPKLDKGRLLTAPKFRDALAGLMVTKSQDHSIAYVPR
FT AAEQLARRASKAPRNPPEARLRLNAGTDEQHPDSQGRITLSADHRYASLSKDCV
FT VIGADYLEIWDQAQWNYQOIHEENFSAASDEALGDIF"
FT complement (5278). .5305)
FT /note="28 bp Inverted repeat at the left end of IS6110;"
FT GAGTCTCCGACATCACCGGGCGGTTCA"
FT complement (5278). .6632)
FT /note="IS6110-6, len: 1355 bp. Insertion sequence IS6110."
FT /insertion_seq="IS6110-6"
FT complement (5320). .56360)
FT /note="Rv2167c, (MTCV270.01), len: 346 aa. Probable IS6110
FT transposase. FASTA best: TRA9 MYCTU P19774 putative
FT transposase for insertion sequence (identical)"
FT /transl_table=11
FT /function="REQUIRED FOR THE TRANSPOSITION OF THE INSERTION
FT ELEMENT IS6110."
FT /locus_tag="Rv2167c"
FT /product="PROBABLE TRANSPOSASE"
FT /protein_id="CAA17494.1"
FT /translation="AEALAGORRIKGRDFKRVGLGRPARPASTLITRTFIADHOG
FT HRSGPGLRWGESICTQLTELGVPIAPSTVYDHNREPSRELROGELKEHISRVHAA
FT NYGVYGNARKVWLTLNREGIEVACTVERLMTKLGLSGTREGKARRTIADPATRAPDL
FT QRRFGPPANRLWADLTIVTWAGFAVYFVDAYARKILGRVASTNATSVLDLAI
FT EQALWTRQQGVLDLKDVIHHTDRGQYTSIRFSERLAEAGIQPSVGAVGSSYDNALAE
FT TINGLYKTELKPGKPMRSIEDVELATARWDFWHRRLYQYCGDVPVPELEAAYQAR
FT QRPAAG"
FT complement (6255). .6581)
FT /note="Rv2168c, (MTV021.01c), len: 108 aa. Probable IS6110
FT transposase. FASTA scores: O08155|O08155 HYPOTHETICAL 12.0
FT kDa PROTEIN (108 aa) opt: 697, E(): 0; (100.0% identity in
FT 108 aa overlap). Tbpase score is 0.928."
FT /transl_table=11
FT /function="REQUIRED FOR THE TRANSPOSITION OF THE INSERTION
FT ELEMENT IS6110."
FT /locus_tag="Rv2168c"
FT /product="PROBABLE TRANSPOSASE"
FT /protein_id="CAA17472.1"
FT /translation="NSGGSRRYPPELRRERAVMVAEIRGQHDSEWALISEVARLILGVG
FT CAVTRKVRQAOVDAGAPGTTTEESAEKRLRRDRAELRRANAILKTASAPFAELD
FT RPAR"
FT complement (6605). .6632)
FT /note="28 bp Inverted repeat at the right end of IS6110,
FT TGAACGCCCGCCGTCGCGAGACTC"
FT complement (6726). .7130)
FT /evidence=EXPERIMENTAL
```

Query Match 95.0%; Score 19; DB 15; Length 348247;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GAGCGCCCGACGCCCAAA 20

|||||  
Db 179298 GAGCGCCCGACGCCCAAA 179316

```

RESULT 4
RMU22370/c      1073 bp  DNA      linear      BCT 24-NOV-1995
LOCUS           Rhizobium melliloti insertion sequence ISRM2011-2 putative
DEFINITION      transposase and ORF genes, complete cds.
ACCESSION       U22370
VERSION         U22370.1  GI:1072026
KEYWORDS
SOURCE          Sinorhizobium melliloti (Rhizobium melliloti)
ORGANISM        Sinorhizobium melliloti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
REFERENCE       1 (bases 1 to 1073)
AUTHORS         Selbitschka,W., Arnold,W., Jording,D., Kosier,B., Toro,N. and
                 Puhler,A.
TITLE           The insertion sequence element ISRM2011-2 belongs to the IS630-Tc1
                 family of transposable elements and is abundant in Rhizobium
                 melliloti
JOURNAL          Gene 163 (1), 59-64 (1995)
MEDLINE         96001244
PUBMED          7557479
REFERENCE       2 (bases 1 to 1073)
AUTHORS         Selbitschka,W.
DIRECT SUBMISSION
TITLE           Submitted (08-MAR-1995) Werner Selbitschka, Dept. of Genetics,
                 University of Bielefeld, Universitaetsstr. 25, Bielefeld, NRW 33501,
                 Germany
FEATURES         Location/Qualifiers
source           1..1073
                 /organism="Sinorhizobium melliloti"
                 /mol_type="genomic DNA"
                 /strain="2011"
                 /db_xref="taxon:382"
source           11..1063
                 /organism="Sinorhizobium melliloti"
                 /mol_type="genomic DNA"
                 /strain="2011"
                 /db_xref="taxon:382"
                 /insertion_seq="ISRM2011-2"
misc_feature     1..10
                 /notes="part of E. coli rpsL gene"
misc_feature     9..10
                 /notes="target duplication"
repeat_unit     11..29
                 /rpt_type="inverted"
RBS             89..94
CDS             join(104..451,451..1050)
                 /function="mediates transposition"
                 /notes="putative transposase"
                 /codon_start=1
                 /transl_table=11
                 /protein_id="AAC43490.1"
                 /db_xref="GI:1072027"
                 /translations="MARFSDLRERVDVAVTGEGLSCFAAKRRGIGISTADWVR
                 FRGTGAAPQMGHPRKLSGLHRAWLCRCRERDFTLHGLVASLSEKGLKVDTRAV
                 WTVHEGLSYKRTIVASEREPDVARHARWLKPCPIDPARLVFIDETWTKNNA
                 PLRWGPRGERLVGYAPFGWNTMTFVAALRADRVSAFFILDGPNGERFRFYVOQVL
                 VPLKAGDVIILNIGSHKGQEIIRAAIRKAGARLFLPKYSPDLNPIELFAKIKHL
                 REAQRSDAIHDLRHILQAVTPQCAAYFEKAGYERA"
                 complement(439..810)
                 /functions="unknown"
                 /notes="ORF"
                 /codon_start=1
                 /transl_table=11
                 /protein_id="AAC43491.1"
                 /db_xref="GI:1072028"
                 /translations="MGAEIVQDDVAGFEFRHNLDDIDAEAFADRAIEDKGSADAV
                 GPFCGKGGHGVPAEGVAGVHQSPAARRPAPQRHVRPLRLGLIDENAGGINSGLVLP
                 SCPVAGDVGLPLFAGDQRULF"
misc_feature     443..450
                 /notes="putative ribosomal slippage site"
stem_loop      452..466
                 /notes="inverted repeat"

```

```

/function="putative involvement in ribosomal
frameshifting"
repeat_unit     1045..1063
                 /rpt_type="inverted"
misc_feature     1064..1073
                 /notes="part of E. coli rpsL gene"
misc_feature     1084..1085
                 /notes="target duplication"

ORIGIN
Query Match      87.0%; Score 17.4; DB 1; Length 1073;
Best Local Similarity 94.7%; Pred. No. 9.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CGAGGCGCCCGAGCCCA 19
        |||||
DB      605 CGAGGCGCCCGAGCCCA 587

RESULT 5
AK098077/c
LOCUS           AK098077      2253 bp      mRNA      linear      PRI 15-JUL-2002
DEFINITION      Homo sapiens cDNA FLJ40758 fis, clone TRACH2002333, highly similar
                 to Human phosphotyrosine independent ligand p62 for the Lck SH2
                 domain mRNA.
ACCESSION       AK098077
VERSION         AK098077.1  GI:21758012
KEYWORDS        oligo capping; fis (full insert sequence).
SOURCE          Homo sapiens
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1
AUTHORS         Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
                 Arita,M., Musashino,K., Tsuki,H., Hara,H., Sugiyama,T., Irie,R.,
                 Osuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J.,
                 Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K.,
                 Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H.,
                 Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A.,
                 Kawakami,B., Suzuki,Y., Sugano,S., Nagahara,K., Masuho,Y., Nagai,K.
                 and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
REFERENCE       2 (bases 1 to 2253)
AUTHORS         Isogai,T. and Yamamoto,J.
TITLE           Direct Submission
JOURNAL          Submitted (04-JUL-2002) Takao Isogai, FLJ project (HRI Team); 2-6-7
                 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
                 (E-mail:genomics@hri.co.jp, Tel.81-438-52-3975, Fax:81-438-52-3986)
COMMENT          NEDO human cDNA sequencing project supported by Ministry of
                 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
                 Research Association for Biotechnology (RAB); cDNA library
                 construction: Helix Research Institute (HRI) (supported by Japan
                 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
                 HRI, and Biotechnology Center, National Institute of Technology and
                 Evaluation; clone selection for full insert sequencing: HRI and
                 RAB; annotation: HRI and RAB.
                 Location/Qualifiers
FEATURES         1..2253
                 /organism="Homo sapiens"
                 /mol_type="mRNA"
                 /db_xref="taxon:9606"
                 /clone="TRACH2002333"
                 /tissue_type="trachea"
                 /clone_lib="TRACH2"
                 /note="cloning vector: pME18SFL3"

ORIGIN
Query Match      87.0%; Score 17.4; DB 9; Length 2253;
Best Local Similarity 94.7%; Pred. No. 9.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GAGGCGCCCGAGCCCA 20

```

```

Db      89 GAGGGCCAGCCACCA 71
|||||
RESULT 6
AE007196/c
LOCUS   Sinorhizobium meliloti plasmid pSymA section 2 of 121 of the
DEFINITION complete plasmid sequence.
ACCESSION AE007196
VERSION   AE006469
KEYWORDS
SOURCE   Sinorhizobium meliloti (Rhizobium meliloti)
ORGANISM Sinorhizobium meliloti
REFERENCE
AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kallman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
TITLE    Nucleotide sequence and predicted functions of the entire
Sinorhizobium meliloti pSymA megaplasmid
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)
MEDLINE 21396509
PUBMED   11481432
REFERENCE
AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kallman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
TITLE    Direct Submission
JOURNAL Submitted (29-MAR-2001) Biological Sciences, Stanford University,
371 Serra Mall, Stanford, CA 94305, USA
FEATURES
source
1..10029
/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
/plasmid="pSymA"
158..715
/gene="Sma0017"
158..715
/gene="Sma0017"
/gene="Sma0017"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK6466.1"
/db_xref="GI:14523064"
/translation="MAAADGRRNGGLRQKARRGLPOARRNHLSTTRIGKROAQL
GSGGQPNRGGCCMPLGREIPSEGGGSRKRSAGRRQKMSLTTFIAELIR
AANVERLTPEVEISLDRSVDTIMRRQGTGAASHEARDVIDLQLASARADLSA
RETRVLDADADIRTKLVDLGE"
repeat_region
1310..2368
/standard_name="ISRM2011-2/ISRM11"
/note="SMA3000; predicted by homology"
/rpt_family="ISRM2011-2/ISRM11"
1406..1813
/gene="Sma0018"
1406..1813
/gene="Sma0018"
/function="Elements of external origin; Transposon-related
functions"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Rm2011-2a transposase"
/protein_id="AAK6466.1"
gene
CDS
/db_xref="GI:14523065"
/translation="MARPFSDLRERVDVAVTGEGLSCRAAAKRFQIGISTAIQWRRR
FRETSAAPQOMGHKPRKLSGPHRAWLCRCRERDFTLHGLVAELSERGLKVDYRAV
WTFVHEGLSYKKRWSPANGSGTSPAGHDG"
1879..2352
/gene="Sma0020"
1879..2352
/gene="Sma0020"
/function="Elements of external origin; Transposon-related
functions"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Tm2011-2b transposase"
/protein_id="AAK6466.1"
/db_xref="GI:14523066"
/translation="MAPLSGMAPRGELVGYAPFGHWTMTFVAALRADRVSAFTLD
GPNGERFRIYQOVLVPELKAGDIVILNLSHKQCEIRAAIRKAGARLFLPKYSP
DLMPIELKFAKIGWUREAQARSRAIHDELRLHQAQVTPQCAAFKEAGYERA"
complement(3102..3746)
/gene="Sma0021"
complement(3102..3746)
/gene="Sma0021"
/function="Elements of external origin; Transposon-related
functions"
/note="glimmer prediction, with response regulator
receiver domain"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK6466.1"
/db_xref="GI:14523067"
/translation="MYRLYNHAIARFSLDFGLTGDKTCRLKAQAWLIAELHQLGYR
DMQKRCAYRGLRYSRIKAEPRSEQTSCEADMDPATILLAEALLLDYGA
LADAGFVAVAGVGAKEIVWSADSEVAGVATDIRFELPNGSVARVARETPGIP
IVGTGGALEWRSRGVANSILLEKPFALAQVAVSELLNEPVLISVAPDNP"
4368..4459
/gene="selC"
/note="synonym: Sma0022; selenocysteine tRNA in RdH
cluster"
4368..4459
/gene="selC"
/product="tRNA-Sec"
complement(4903..6030)
/gene="Sma0025"
complement(4903..6030)
/gene="Sma0025"
/function="Miscellaneous; Hypothetical/Global homology"
/note="glimmer prediction, C-terminus similar to
hypothetical protein HI1048 - Haemophilus influenzae
(strain Rd Kw20) with transglutaminase-like superfamily
domain"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK6467.1"
/db_xref="GI:14523068"
/translation="WVILSERAFSRDQTAELVRLNRRLTAATAAATLWVSKG
CMQAREAGNSPWRPRLITIKVDLPQAGPAQLPVVRSAGDYQKADAPLVNSN
TDIRNEKPSGADIILRWMEDEPRVIEVQVATRDRAVRSKIATRELQHELNG
TPSPVGVIVKTTAMRIVSGRDAERARAYIDWIDNTFRDANIDCGGVGNARDMLE
TGYFGKCADISSLFVSLARAGLPARDVFGIRVADSADPKSLGRSGDITKAQCHRAE
VYLDAGWVPVDPADVRKVLLENPLDNPVAPFAFREKAYGNWENWVGYNTARDLVL
PGGELRQGLFMPYPAVTSRGELDLNPNQTFAYSITSREITA"
complement(6076..6504)
/gene="Sma0026"
complement(6076..6504)
/gene="Sma0026"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

```

/protein\_id="AAK64671.1"  
/db\_xref="GI:14523069"  
/translation="MEQAPFNIVGGDEKEIAMSNLAPTRPACVLAAPFLJIALANP  
ASAQNTWEDCAAGAAKAPATPPDRASADGTAPCNSGSGTGTGCGAHGINTPQG  
ATAGTTPQPTTARGIDILATPEIPAPETAAPGTQTPTDC"  
complement (6517..7566)  
/gene="seld"  
/notes="synonym: Sma0028"  
complement (6517..7566)  
/genes="seld"  
/EC\_number="2.7.9.3"  
/function="Small molecule metabolism; Global functions"  
/notes="glimmer prediction"  
/codon\_start=1  
/transl\_table=11  
/product="selenide, water dixinase (selenophosphate  
synthetase), Seld"  
/protein\_id="AAK64672.1"  
/db\_xref="GI:14523070"  
/translation="WNTLAPRLTLAHGGCGCKLAPSLQOLLANPAARPEAQLLV  
GNDMADDAWVDNTCVIATDTPMPVDDPRDGRILATNAISDVYKGGKPLA  
LAIVGPNKLUSTIAKLEGASICAEPVAGCHSIDSVETIYGLAVIGLCHPS  
EVRNGGKAGDALILTKGIGVYSAAIKKALPPGCEEMIGSTILLNIGADLAA  
DPDVALTVTFGLVGHLEMARASQLSLRLSAIPLSQAYMLAQFGITGASGR  
NWASYGADVPLDPLDMQRLLLADPQTSGLLVACAPKAGALVEKRLAGYPLAGI  
VGTAAAGAAQIKVIS"  
complement (7861..8088)  
/gene="Sma0031"  
complement (7861..8088)  
/genes="Sma0031"  
/function="Miscellaneous; Unknown"  
/notes="glimmer prediction"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK64673.1"  
/db\_xref="GI:14523071"  
/translation="MSEWIDFRWDPDKMERPGIVFEVTNGDQTLTGCVVPLPLPS  
DWVHPLRPRAVPQPRRHSSPLPKPAGPQQ"  
complement (8105..9760)  
/genes="Sma0034"  
complement (8105..9760)  
/gene="Sma0034"  
/EC\_number="3.4.24.40"  
/function="Macromolecule metabolism; Macromolecule  
degradation; degradation of proteins, peptides,  
glycopeptides"  
/notes="glimmer prediction; partially similar to serralysin  
(Serratia protease) precursor"  
/codon\_start=1  
/transl\_table=11  
/product="possible protease"  
/protein\_id="AAK64674.1"  
/db\_xref="GI:14523072"  
/translation="WPAATYATTGNAVIGLGDWKXIKDFTFSPPTSASFYSAGY  
GNGEPLKGPAAISGQAATRAALQDFSSVANVTETESATKHADRLASSDAPST  
AWAYFSTAEAGDAWFKSSGHSYRPNVGNAYVTFLETHGALGLBHAEGNMPV  
NRDSMEYTVMSYRSYVGAATTTGTNETWYGAQSLMMYDIALQHMYGADTFHSENT  
TYRSPSTGVNGVNGQAGPGCKILLTVWDGGTDTYSNTYTKLVPLRPGENT  
TTTSAQLAKLHYDGGKVAIGNALNALYQGDTRSLIENAKGAGGNDAITGNAANALW

Query Match 87.0%; Score 17.4; DB 1; Length 10029;  
Best Local Similarity 94.7%; Pred. No. 1e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 CGAGGCGCCAGCCGCCA 19  
Db 1907 CGAGGCGCCAGCCGCCA 1889  
RESULT 7  
AE007297/c  
LOCUS

DEFINITION Sinorhizobium meliloti plasmid pSma section 103 of 121 of the  
complete plasmid sequence.  
ACCESSION AE007297 AE006469  
VERSION AE007297.1 GI:14524266  
KEYWORDS  
SOURCE Sinorhizobium meliloti (Rhizobium meliloti)  
ORGANISM  
REFERENCE  
AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,  
Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,  
Gurjal,M., Hong,A., Huizar,L., Hymen,R.W., Kahn,D., Kahn,M.L.,  
Kalman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,  
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.  
Nucleotide sequence and predicted functions of the entire  
Sinorhizobium meliloti pSma megaplasmid  
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)  
21396509  
11481432  
2 (bases 1 to 10296)  
Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,  
Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,  
Gurjal,M., Hong,A., Huizar,L., Hymen,R.W., Kahn,D., Kahn,M.L.,  
Kalman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,  
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.  
Direct Submission  
Submitted (29-MAR-2001) Biological Sciences, Stanford University,  
371 Serra Mall, Stanford, CA 94305, USA  
FEATURES  
source  
1..10296  
/location/Qualifiers  
/organism="Sinorhizobium meliloti"  
/mol\_type="genomic DNA"  
/strain="1021"  
/db\_xref="taxon:382"  
/plasmid="pSma"  
164..703  
/genes="Sma2029"  
164..703  
/genes="Sma2029"  
/function="Small molecule metabolism"  
/notes="glimmer prediction. Contains similarity to  
Isochorismatase family. Similar to LOCUS 4558196 208 aa  
Chain A, The 1.8 Angstrom Crystal Structure Of The Yeac  
Gene Product From Escherichia Coli Reveals An Octameric  
Hydrolase Of Unknown Specificity. ACCESSION 4558196"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein"  
/protein\_id="AAK65768.1"  
/db\_xref="GI:14524267"  
/translation="MVVNHITLAKLARAFNVPTLTSVIAARGLLFKQITDVPDQ  
EVIDRTVATQDENVNAVKATGRKQIAGLWTEVCVAMPVIAQSGWDTVID  
ASGCIKSGHEVAIQRAAGANVTVMALAGEWQDARTHEVELTILIQHFGS  
GIAYLEOQLLNTVPVSEG"  
714..1085  
/genes="Sma2031"  
714..1085  
/genes="Sma2031"  
/function="Cell Processes; Protection responses;  
detoxification"  
/notes="glimmer prediction. Local similarity to 278 aa  
NON-HEME CHLOROPEXIDASE. Possible protein fragment."  
/codon\_start=1  
/transl\_table=11  
/product="Putative non-heme chloroperoxidase"  
/protein\_id="AAK65769.1"  
/db\_xref="GI:14524268"  
/translation="MAKAVLVAIPPLMLKNDDDPGTPEVDFGFRALAGNRAOFF  
RDVPSGPFSGFNREGAAVHEGVLNQWVQGMKQCALRWQGLLGNRPFWTSRRS  
PCQRRCCTVTRTSFRSPRL"  
923..1222  
/genes="Sma2032"  
gene  
CDS  
gene  
CDS  
gene  
CDS

```

CDS
923..1222
/gene="Sma2032"
/function="Cell Processes; Protection responses; detoxification"
/note="glimmer prediction. Local similarity to 278 aa NON-HEME CHLOROPEROXIDASE. Possible protein fragment."
/codon_start=1
/transl_table=11
/product="Putative non-heme chloroperoxidase"
/protein_id="AAK65770.1"
/db_xref="GI:14524269"
/transl_table="MAAGHGEAPRRITMASRPRKPTKTDLKAITVPTPVLHGEND
OWPIAAGALKVLLPNSGLKTPFGSHGLMTVNADVLNADLLEFITNRSDCGA"
1755..2552
/gene="Sma2033"
/feature="Miscellaneous; Unknown"
/function="Miscellaneous; Unknown"
/note="glimmer prediction. C-terminus is similar to LOCUS BAA19632 376 aa lytic transglycosylase [Actinobacillus actinomycetemcomitans]. ACCESSION BAA19632"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65771.1"
/db_xref="GI:14524270"
/transl_table="MAFFAKQLIAAGSLAFCLLTVCVAEASCGKGAAGFOQWKVEF
AETAKAGVRGKGLAALIGAXVGTIRADRSVNKAPSGVDDFWRRGSAIISKGR
SLKTAALALFKIERIDVPAGVLLAIMEIGTIGGASMGNTVSAITVLYADGERQ
FPAPHAIALKLVDSGLSARSVAGHMEIGTGPFGNVLKYGVRNMRDSTALM
STANFLKAGWRAGGYGQNMGAITAGWNSASVYQKAIARIGEAIDGR"
complement(3483..3728)
/gene="Sma2034"
/complement(3483..3728)
/gene="Sma2034"
/function="Miscellaneous; Unknown"
/note="glimmer prediction. Similar to C-terminus of LOCUS BAA19632 376 aa lytic transglycosylase [Actinobacillus actinomycetemcomitans]. ACCESSION BAA19632"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65772.1"
/db_xref="GI:14524271"
/transl_table="WVEVTLWGSLAAAGNSKVEIAKNIRELFARLSERFPRLEPL
MARGIAVADGTVYRDTWSKELPTGAEIYLLPRLAGG"
complement(3738..6170)
/gene="Sma2037"
/complement(3738..6170)
/gene="Sma2037"
/function="Small molecule metabolism"
/note="glimmer prediction. Similar to motifs found in aldehyde oxidase and xanthine dehydrogenase, C terminus. Similar to LOCUS CAA05038 769 aa 4-Hydroxybenzoyl-CoA reductase alpha-subunit [Thaurea aromatica]. ACCESSION CAA05038"
/codon_start=1
/transl_table=11
/product="Putative oxidoreductase"
/protein_id="AAK65773.1"
/db_xref="GI:14524272"
/transl_table="WVHLTRLPCHRCKSVAEESGSDRRGNLTLCRKSLSIHLRQD
RARRGRSRHSHATIRKIDTLKAQLPGKAVLTAADLPDITDGDAMYLINCMGA
QIVGRILRSPHAIIRKIDTLKAQLPGKAVLTAADLPDITDGDAMYLINCMGA
RKALYDGHAVAAVAAIARTAKALIEVEYELLPHVTDVQAFAPDLINDTIF
TTGVDPKPRNSMRSQFGHGDVDMARADVFVETFKTEOHOYIPSPHACVAS
VSSDGTADLVCTQGHVYRCHCAOLLGMDASKLAVTSSEIGGFGGKTHVWAEFVAL
ALSRKGRPVKLMTDEVRASGTSATSIDVKIGALKDGTILAADATLAYSQGPYA
GAAEVMGATFACYLENRTVGEVLNVKTAAYRAPAPAAFAVEASIDELAX
KYVMNPIDRIKNAQEGTKASGYPIGIGPTLEAVKNHPHMKPLGRNQGRGA
CGFWFNFGQCTDLNIGMDGTVSLAVGTVDVGSGRASLSLVAAEBELGIDYSQRTVI"

gene
complement(6469..7329)
/gene="Sma2041"
/complement(6469..7329)
/gene="Sma2041"
/function="Small molecule metabolism"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="probable oxidoreductase"
/protein_id="AAK65774.1"
/db_xref="GI:14524273"
/transl_table="MRYIRPSIEDAVGLLAASGKAAVLAGSDLLVRMKGFIPEP
LIIDIKAIADLRHITESDEDFVIGAAVPCAVLGENAALRRAPGVVEGANLIGSKQYQ
GRTIYVNCNAPSAADSVFALLAGAEALIRGPAAGSTIAYEMVPGVKISLAPGE
II EAILDKRQFSDAYQRTPTREMDIAYVSAVVKLTLDQGVQVSARVALGAAP
TVLLVEAAEILTGSRVDDKTLDRLAAGSGACRPIDDKRGTVFERRKVAGVLAKRVA
LAAVQRAGLE"
7882..8941
/standard_name="ISRM2011/ISRM11"
/note="Sma3016; predicted by homology"
/evidence=not_experimental
/rpt_family="ISRM2011-2/ ISRM11"
7977..8387
/gene="Sma2043"
/feature="Miscellaneous; Unknown"
/function="Elements of external origin; Transposon-related functions"
/codon_start=1
/transl_table=11
/product="TSM2011-2a transposase"
/protein_id="AAK65775.1"
/db_xref="GI:14524274"
/transl_table="MARPPSNDLRERVDAVTGEGLSGCRAAAKRFGHHRDLGAA
VSRDQGRTPDQCAQPAQAFRAGLAALPLPRALHAARTCRRVARPEGGFICA
WTFVHEEGLSYKRRWSPANGSGTSPATHDG"
8453..8926
/gene="Sma2045"
/feature="Miscellaneous; Unknown"
/function="Elements of external origin; Transposon-related functions"
/codon_start=1
/transl_table=11
/product="TSM2011-2b transposase"
/protein_id="AAK65776.1"
/db_xref="GI:14524275"
/transl_table="MAPLRGWRPGRVLGYAFGHWNWTFVAALRADRVSAFFILD"

Query Match 87.0%; Score 17.4; DB 1; Length 10296;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCA 19
Db 8481 CGAGGCGCCAGCCCA 8463

RESULT 8
AE007314/c
LOCUS AE007314 10578 bp DNA linear BCT 15-AUG-2001
DEFINITION Sinorhizobium meliloti plasmid pSyma section 120 of 121 of the complete plasmid sequence.
ACCESSION AE007314 AE006469
VERSION AE007314.1 GI:14524456
KEYWORDS Sinorhizobium meliloti (Rhizobium meliloti)
ORGANISM Sinorhizobium meliloti

```

Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.

REFERENCE 1 (bases 1 to 10578)  
 AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P., Barloy-Hubier,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J., Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L., Kalman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R., Wells,D.H., Yeh,K.-C., Davis,R.W., Pederspiel,N.A. and Long,S.R.

TITLE Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti pSymA megaplasmid

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)

FEATURES

source 1.10578  
 Location/Qualifiers  
 /organism="Sinorhizobium meliloti"  
 /mol\_type="genomic DNA"  
 /strain="1021"  
 /db\_xref="taxon:382"  
 /plasmid="pSymA"  
 420..1478  
 /standard\_name="ISRM2011-2/ISRM11"  
 /notes="Sma3018; predicted by homology"  
 /evidence="not experimental"  
 /rpt\_family="ISRM2011-2/ISRM11"  
 516..923  
 /gene="Sma2373"  
 516..923  
 /gene="Sma2373"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Tm2011-2a transposase"  
 /protein\_id="AAK65941.1"  
 /db\_xref="GI:14524457"  
 /translation="MARFSDLRVVDVAVTGBGLSCRAAKRFGIGISTADWRR  
 FRTSGAPQGGKPKRLSGPHRWLLCRERDFTLHGLVABLSEGLKVDYRAV  
 WTVFEEGLSYKRRWSPANGSGPTSPATGHDG"  
 989..1462  
 /gene="Sma2375"  
 989..1462  
 /gene="Sma2375"  
 /functions="Elements of external origin; Transposon-related functions"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Tm2011-2b transposase"  
 /protein\_id="AAK65942.1"  
 /db\_xref="GI:14524458"  
 /translation="WAPLRGWAPRGERLVGAPFCHWNTWTFVVALRADRVSAFILD  
 GPINGFRFRIYQVGLVPELKGDIVLDNGLSHKGQEIARAIRKAGARLFFLPKISP  
 DLNPIKLFPAKIKHMLREQAQRSDAIHDELHILQAVTPQCAAYFKEAGYERA"  
 1664..2833  
 /gene="Sma2377"  
 1664..2833  
 /gene="Sma2377"  
 /function="Cell processes; Transport of small molecules"  
 /notes="glimmer prediction; similar to putative chloramphenicol transporter in several organisms, including Salmonella typhimurium, AF118107, Klebsiella pneumoniae, AAF27726, and Enterobacter aerogenes, AAD22144"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative transmembrane transport protein"

gene  
 CDS  
 /protein\_id="AAK65943.1"  
 /db\_xref="GI:14524459"  
 /translation="MTVNVSPVTVMFLISFTVLGIAGTDLVLPALPAMPALTGGTAA  
 LAQVLAAYAGGTLVGLTFEGELGARYSRKLLVMSLGLFAVTSLSASVAPTLEWLV  
 LRPAQAGSGPAVFAPGLFGLYPGDKAPSMFGRGCSIESLTPALGAPAGVMTVG  
 WQTSFLMLAGLAILCAVGSWAYRQSLPRLALEVQSVMSIIRNGDFLRHGLSOAL  
 SLGSLIFVGPAPAVMTGLMTGTFILLOVFGIALFILASNASNALPHOAIVASGRBI  
 MITSGLIFVGLLILVTSLSGRSLVPLVPLANTANGAGIRGPIGPHOAIVASGRBI  
 SRGALVVAAILGITAGTARARAFINVGWPLALSSALLALLLCLKLGSTA"  
 complement(2924..5107)  
 /gene="Sma2379"  
 complement(2924..5107)  
 /gene="Sma2379"  
 /EC\_number="1.11.1.6"  
 /function="Cell Processes; Protection responses;  
 detoxification"  
 /notes="glimmer prediction; highly similar to  
 catalase/peroxidase from Streptomyces reticuli, CAA74698"  
 /codon\_start=1  
 /transl\_table=11  
 /product="catalase/peroxidase"  
 /protein\_id="AAK65944.1"  
 /db\_xref="GI:14524460"  
 /translation="MDQKSDSAGKCPVAHTAPGRSNRDNWPDQLDVQVLRHSGLSLSD  
 PLGNTFYAEFFKLLDLDALRDLRALMTDSQWPFADFGHGLGFIRMAWHSAGTYR  
 ITDGRGAGQQRFPAPLNSPDNANLDKARLLWPKQYGNRISSWADLLLTGNVA  
 LESMGFKTFAGGRVDMVPELFWGPGTGLGDERYSGERQLSPALAAVQMLTYV  
 NPEPGNPDVAAARDIRETFARMAANDEETVALLIAGGTFGKTHGAGDPGFIGADP  
 EGGAIEDGLGWKSTFGTGKDAITGGPEVTSOTPTKSNHFFENLHNEWELTKS  
 PAGAYQWAKNAEATIPDAYDPSRKHVPTLTDLSLRDPAYEKLSRSLFENPDFA  
 DAFARWFKLTHRDMPKRVLPEDLIWQDVPVADKRLVDETIDIAKAKII  
 ASGLSVQELVSTASASTFRGSDKRGAGNARIRLAPQKDMVNPQALRLVSLYE  
 GIORDFNAQTDGKISLADILVLAGAAVAAKAGHDITVPTFGMSDASEAQT  
 AASFAALEPRADGRFNVVSTTROQFMKPEALVDRALLLTAPETVLTGGLRLVKA  
 GEPKHGVTSRPEALTNDFVLLDNGTOWSPIEGEGVVEGREDRTGAARTGTTRVD  
 LIIFGSHSQARAFVAYQAQSDAREKFKVDFAVAATKYNADRPDLV"  
 5640..6113  
 /gene="Sma2381"  
 5640..6113  
 /gene="Sma2381"  
 /function="Miscellaneous; Unknown"  
 /notes="glimmer prediction"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAK65945.1"  
 /db\_xref="GI:14524461"  
 /translation="MLLKAMKIGRQSLDYTFGDKWLYCLAVERYGASSETSALHALLR  
 GKPRALEGIRLVNVRVVDNANOACLVNSICEFGQSRPDLAALHHAADRRLKNVVER  
 IREQAAGDLAVSLPAEAVADFLVANIAGIRIARGGARREHLQSLRLALRAVT"  
 6167..7135  
 /gene="Sma2383"  
 6167..7135  
 /gene="Sma2383"  
 /function="Small molecule metabolism"  
 /notes="glimmer prediction; similar to TED2, a putative  
 quinone oxidoreductase, from Zinnia elegans, BAA06460, and  
 from cowpea, T11672; also similar to putative  
 oxidoreductases from Mycobacterium tuberculosis,  
 Agrobacterium tumefaciens, etc. Zinc-binding family"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable oxidoreductase"  
 /protein\_id="AAK65946.1"  
 /db\_xref="GI:14524462"  
 /translation="MAKAVNKEVGCTDVMVEFVDRPEFVAPRHVVVEVAAAGVNFMDI  
 GVFGQWATDIPNPKVLVGEAGRVLAGDGTGEFAGVGRVAVVYAPGSAQRQSLPA  
 ASLVKIPDVTDDRTAAATMQGLTASHFATDFYVQPGDIALVHAAGGVGLLTGII  
 RLKRGVIGRVSSSDKVAIARKAGASHVITDTRGFADVLRLTGGGVNVYDGGSP  
 KTFKGSIALRRSGTFCFVGLGPGPLEIMNLPKSIKIGYATFMDFHITRELLDR  
 TKOLFWDIEDGSIITVIGTYRLADAAGAHAAMASRATTGKLLIIP"  
 complement(7357..8946)  
 /gene="Sma2385"



```

CDS
    complement(7357..8946)
    /gene="Sma2385"
    /function="Cell processes; Transport of small molecules"
    /note="glimmer prediction; similar to probable ABC
    transporter from Streptomyces coelicolor, T35245, and to
    VarM (probable transporter for virginiamycin) from
    Streptomyces virginiae, BAA96297. ABC transporter family"
    /codon_start=1
    /transl_table=11
    /product="probable ABC transporter, ATP-binding protein"
    /protein_id="AAK65947.1"
    /db_xref="GI:14524463"
    /translation="MPSITLIALSKSPDGEHVSFSLDLAPGERTGLVGRNGIGKTS
    VLNIAIGLRSSGTVAIQGRVALRAQLRAGADETTADVFCATQAVAVLRARAKGDA
    SVELETADWTVERIVSALARGLEARDLILNOLSGGQRTAVLAAVLRSEPDFLL
    LDEPTNNIDREARVIGLLSGWRSAGTAVVSHRELEEMDAIIELSLGTRKGGWM
    SAYQAARVELEAAQOQSLTARKTADVEDRKALAEILDKRDAAGTGRKAAGMPRI
    LVGRKKSVAELSRGSLAEERRAGALDAVTAARKEVLQPFSLRTELPELPAQRQ
    VLAQGVTAQDPARPIRLDSFSLVGRPRVSVTPGNSGKTSLLKVTGLPFPKGT
    VSNVVPFTLLQOSVLSRGETLLENFKELNPGASDNACRAALASFRADAALQRYE
    ALSGGQVRLAGLACALGSDPSPSLIIDEPTNHLIDISIEAVEAGLLSYDGLVTVSH
    DETFLANIGIGTRVELSTSRG"
    gene
    9057..9713
    /gene="Sma2387"
    CDS
    9057..9713
    /gene="Sma2387"
    /function="Miscellaneous; Unknown"
    /note="glimmer prediction; putative bacterial regulatory
    protein, tetr family"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAK65948.1"
    /db_xref="GI:14524464"
    /translation="MIRARRAGRPTRREAEALTRLLDSARSTFARKGIANSMEE
    IAAELGSKHLYRYPNRQALLAVRDLVRKTLAEAGGGEAPLAALRDMAR
    YFRFGTDYSAFLSVTAESAFLPLERLAAWSSAALEPLVQAIISAQAAGLVVFG
    STIETCHVLIDLEGANVRVLCLESPPDASERLRFESRWAVFQTAMMPENRPLGV
    "
    gene
    complement(9922..10455)
    /gene="Sma2389"
    CDS
    complement(9922..10455)
    /gene="Sma2389"
    /function="Cell processes; Adaptation"
    /note="glimmer prediction; very similar to Ohr, a putative
    stress-induced protein, from Bradyrhizobium japonicum,
    Query Match 87.0%; Score 17.4; DB 1; Length 10578;
    Best Local Similarity 94.7%; Pred. No. 1e+03;
    Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCGACGCCACCA 19
|||||
Db 1017 CGAGGCGCCGACGCCGCA 999

RESULT 9
AE007285
LOCUS
DEFINITION
Sinorhizobium meliloti plasmid pSymA section 91 of 121 of the
complete plasmid sequence.
ACCESSION
AE007285 AB006469
VERSION
AE007285.1 GI:14524132
KEYWORDS
Sinorhizobium meliloti (Rhizobium meliloti)
SOURCE
Sinorhizobium meliloti
Organism: Sinorhizobium meliloti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
1 (bases 1 to 10931)
REFERENCE
Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barloy-Hubier,F., Bowser,L., Capela,D., Gallibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kalman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
Direct Submission
Submitted (29-MAR-2001) Biological Sciences, Stanford University,
371 Serra Mall, Stanford, CA 94305, USA
FEATURES
source
1..10931
/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
/plasmid="pSymA"
/complement(218..568)
/gene="Sma1791"
/complement(218..568)
/gene="Sma1791"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65646.1"
/db_xref="GI:14524133"
/translation="MDVGLPCITSSPPGCGSRAAATAGNVPRRLSTRVAARCCRP
QTPSAVRATEPDDDDVHLSELHENHSACHMINVLFNVRKVVIDLENILDLRRSFH
NTRPVPFVAPRSVA"
962..1384
/gene="Sma1793"
962..1384
/gene="Sma1793"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65647.1"
/db_xref="GI:14524134"
/translation="MEAVNRRSTLALGLTMAATPLIAWVTPAAATYGPDEGEIGPG
VRVVALGERASVTPAYKAVKLRDVVIOGAKTPTDNVMTDMLCHMTGELSVYQNEKK
FTVKGDVWTCAKADITTEGTQNTSNTSVAINRIIDLMTS"
2609..2935
/gene="Sma1797"
2609..2935
/gene="Sma1797"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65648.1"
/db_xref="GI:14524135"
/translation="MSPKSRHQPQAYEPACGGIMADAOKMKKLIEDCERDIAACLHPE
SGIPEHLIQQLLSLRLDQQAKEALGDWQGRWHDPEGDDMSPSPLWWTQPEFFGT
IWRLLKI"
3551..5452
/gene="kup2"
3551..5452
/note="synonym: Sma1798"
/gene="kup2"
/function="Cell processes; Transport of small molecules;
cations"
/note="Similar to LOCUS KUP_ECOLI a 622 aa which encodes
the KUP SYSTEM POTASSIUM UPTAKE PROTEIN. ACCESSION

```

```

TITLE
Nucleotide sequence and predicted functions of the entire
Sinorhizobium meliloti pSymA megaplasmid
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)
MEDLINE
21396509
PUBMED
11481432
REFERENCE
2 (bases 1 to 10931)
AUTHORS
Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barloy-Hubier,F., Bowser,L., Capela,D., Gallibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kalman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
Direct Submission
Submitted (29-MAR-2001) Biological Sciences, Stanford University,
371 Serra Mall, Stanford, CA 94305, USA
FEATURES
source
1..10931
/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
/plasmid="pSymA"
/complement(218..568)
/gene="Sma1791"
/complement(218..568)
/gene="Sma1791"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65646.1"
/db_xref="GI:14524133"
/translation="MDVGLPCITSSPPGCGSRAAATAGNVPRRLSTRVAARCCRP
QTPSAVRATEPDDDDVHLSELHENHSACHMINVLFNVRKVVIDLENILDLRRSFH
NTRPVPFVAPRSVA"
962..1384
/gene="Sma1793"
962..1384
/gene="Sma1793"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65647.1"
/db_xref="GI:14524134"
/translation="MEAVNRRSTLALGLTMAATPLIAWVTPAAATYGPDEGEIGPG
VRVVALGERASVTPAYKAVKLRDVVIOGAKTPTDNVMTDMLCHMTGELSVYQNEKK
FTVKGDVWTCAKADITTEGTQNTSNTSVAINRIIDLMTS"
2609..2935
/gene="Sma1797"
2609..2935
/gene="Sma1797"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65648.1"
/db_xref="GI:14524135"
/translation="MSPKSRHQPQAYEPACGGIMADAOKMKKLIEDCERDIAACLHPE
SGIPEHLIQQLLSLRLDQQAKEALGDWQGRWHDPEGDDMSPSPLWWTQPEFFGT
IWRLLKI"
3551..5452
/gene="kup2"
3551..5452
/note="synonym: Sma1798"
/gene="kup2"
/function="Cell processes; Transport of small molecules;
cations"
/note="Similar to LOCUS KUP_ECOLI a 622 aa which encodes
the KUP SYSTEM POTASSIUM UPTAKE PROTEIN. ACCESSION

```

```

P30016. "
/codon_start=1
/transl_table=11
/product="Kup2 Potassium uptake protein"
/protein_id="AAK65649.1"
/db_xref="GI:14524136"
/translation="WADSLDHAPQAQNNLPQFLATIGAVGVYGDIGTSLPYAPREA
LSPGPGVGRDEVIGLSLWLTATVITKYVLFLLRDNDSGGTSLSLALLKK
GTYKPYLMPFAGVLGNALEIGDAMITPALSIVLSAVEGLKLVAPALHDVLPISVIL
LUFVDSRGTAIVFVFGITLIVFLMAAGVAGHIGDAILLSAFNPLNIGFLWNA
GLIGFTVGLAITVTTGALYDILGHFRHSIQANFAVFPALNALYLGQALVLS
HPDAINPPFLFPNFWALLPMVILATAATIASQVITCAFSILRQALHGLFLRPEI
CYTSETQIQVLPVITLTVLALMLFSGSEALAPAYGVITGAMVDTILFAFE
FYVRGWPALTAIAVLPLFNLVFLCANLLKVHGGYVPIIAGTILITMMWTWRK
GVSVREKTARODIPLSQMAIVERKSHAPVQVPTAFELTATPTDTPPAVLLNIKH
NVLHQNVMITIKTARVYVPEKORYTITKLSDFSLLELRFEGMDQNVSRALRC
REGKFEIMSTSYFUGRKLLADFPQGLPQWQDKLFAMADSAIDPTEYFHLPNRV
VELGEQV11"
/complement(6078..6583)
/standard_names="ISRM2011-2; ISRM11"
/notes="Sna3014; fragment of insertion sequence; predicted
by homology"
/evidence="not experimental"
/rpt_family="ISRM2011-2/ISRM11"
/complement(6092..6565)
/genes="Sma1799"
/complement(6092..6565)
/genes="Sma1799"
/functions="elements of external origin; transposon-related
functions"
/codon_start=1
/transl_table=11
/product="TRM2011-2b transposase"
/protein_id="AAK65650.1"
/db_xref="GI:14524137"
/translation="WAPLPGMAPRGERLVGVPFGHWTMTFVAALRADRVSAFPLD
GNGRFRRIYVQVLVPELKGADIVLDNLGSHGQEIARAIRKAGARLEPLKYP
DLNPIELFAKIKHMLREACARSDAIHDELRLHQVTPQECAYFKEAGYERA"
/complement(6661..7920)
/genes="Sma1801"
/complement(6661..7920)
/genes="Sma1801"
/functions="elements of external origin; transposon-related
functions"
/notes="identical (99%) to reverse transcriptase/maturase
[Sinorhizobium meliloti]. ACCESSION CAA72334."
/codon_start=1
/transl_table=11
/product="Reverse transcriptase, maturase"
/protein_id="AAK65651.1"
/db_xref="GI:14524138"
/translation="MTSESTDKPRIEKRRIYVAYKAVKANGAGVGDQLTLEIFEK
DLAANLYKWNRSYSGTIFPPPRAVRIPKKGAGGERVLGVTVDRIAQMVKKMIEP
DLDSLPLPSYGYRPGKSALDAVTVTRQCKYDVLDFDNLPHLLKXAVR
KQVKNWALLYIERMLTAMENKNGVIERSGTPOQGVSPILANFLHYAFDLWMT
TPDLPWCYADDGLVHCSEQOALVELSRLAACGLQHPKTKIVYCKDORR
EAYPNVTFPLGQVPRVANTONDERFCGTPAVSPITALKMEATIKSLNIPTPT
GTLAEIAKQNLPLRGMWYAYGYRSRSLSTLADYVNOKLRAWIRKFKRQSHKTRA
SULPLKARENGLFVHWKAGTNTFT"
/complement(8466..9015)
/standard_names="ISRM2011-2/ISRM11"
/notes="Sna3015; fragment of insertion sequence; predicted
by homology"
/rpt_family="ISRM2011-2/ISRM11"
/complement(8515..8922)
/genes="Sma1803"
/complement(8515..8922)
/genes="Sma1803"
/functions="elements of external origin; transposon-related
functions"
/codon_start=1
/transl_table=11
/product="TRM2011-2a transposase"

```

```

/protein_id="AAK65652.1"
/db_xref="GI:14524139"
/translation="MARPFNDLRRVVDAVTGEGLSCRAAAKRFGIGISTALDWVR
FRETGAAPQMGCHAPKRLSGRHWLLCRNRERDFTLHGLVAELSERGLKVDYRAV
WTFVHEGLSYKKRWSWPANGSGTSPATCHDG"
/9003..9422
/genes="Sma1806"
/9003..9422
/genes="Sma1806"
/function="Miscellaneous; Unknown"
/notes="Glimmer prediction"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65653.1"
/db_xref="GI:14524140"
/translation="MHPALERMQRGKRNLTALAADPVNFTGSELGKFAATTEMQOCD
QRTSQSEBPTVAQRLLIFRAHLPTNRGCAIVASRLRRRAADRLSRTERRRTRWDA
RGRSSAWSAAPRPAYRCFFIAVASASGPCNNLLS"
/complement(9668..10072)
/genes="Sma1808"
/complement(9668..10072)
/genes="Sma1808"
/function="Miscellaneous; Unknown"
/notes="Glimmer prediction. Overlaps in opposite direction
with Sma1807"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65654.1"
/db_xref="GI:14524141"
/translation="MEISFAPQSRRCAPARTRPKTSGLRSWFREGVHKLVDKQTL
GGSLVVGQAQIASVSDASAKATSGDRLRSSATASRRPTPISHRYAASRTYRNA
RPSASCPWCSTSLTRKEHGSALAPIQVT"
/complement(10090..10797)
/genes="Sma1809"
/complement(10090..10797)
/genes="Sma1809"
/EC_number="1.11.1.10"

Query Match      87.0%; Score 17.4; DB 1; Length 10931;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 CGAGGCGCCCGACCCACCA 19
DB 6537 CGAGGCGCCCGACCCCGCA 6555

RESULT 10
LOCUS BTNXGENE 12706 bp mRNA linear MAM 03-MAR-1999
DEFINITION B.taurus mRNA for tenascin-X.
ACCESSION Y11915
VERSION Y11915.1 GI:2462978
KEYWORDS flexilin; tenascin-X; TN-X gene.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 12706)
AUTHORS Elefteriou,F., Exposito,J.Y., Garrone,R. and Lethias,C.
TITLE Characterization of the bovine tenascin-X
JOURNAL J. Biol. Chem. 272 (36), 22866-22874 (1997)
MEDLINE 97426436
PUBMED 9278449
REFERENCE 2 (bases 1 to 12706)
AUTHORS Elefteriou,F.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1997) F. Elefteriou, Institut de Biologie et
Chimie des Proteines, CNRS, UPR-412, 7 passage du Vercors, 69367
Lyon Cedex 07, FRANCE

```

[illegible]

Chromosome 16 libraries constructed by Norman Doggett

(unpublished). VECTOR: sCos-1  
 IMPORTANT: This sequence is not the entire insert of clone  
 LA16-316G12 it may be shorter because we sequence overlapping  
 sections only once, except for a 100 base overlap.  
 The true left end of clone LA16-316G12 is at 1 in this sequence.  
 The true left end of clone LA16-399E4 is at 37377 in this sequence.  
 The true right end of clone LA16-358B7 is at 21289 in this  
 sequence.

#### FEATURES

##### source

1. 37476  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosomes="16"

/clone="LA16-316G12"  
 /clone\_lib="LA16"

##### misc\_feature

987. 2805

/notes="CpG island"

/evidence=not experimental

gene  
 join(1727. 1820,6611. 6751,6973. 7060,7235. 7315,  
 7567. 7674,8993. 9052,9200. 9266,9365. 9560,9833. 9937,  
 10057. 10134,10252. 10354,10598. 10705,10790. 10885,  
 10984. 11087,11464. 11584,12093. 12217,12298. 12388,  
 12540. 12596,12672. 12769,12869. 12920,13040. 13170,  
 13200. 13443,13818. 13961,14057. 14149,14241. 14432,  
 14513. 14581,14676. 14755,14918. 15026,15369. 15440,  
 15616. 15696,15772. 15892,15965. 16088,16167. 16372,  
 16450. 17510)  
 /genes="c316G12.1"

##### mRNA

join(1727. 1820,6611. 6751,6973. 7060,7235. 7315,  
 7567. 7674,8993. 9052,9200. 9266,9365. 9560,9833. 9937,  
 10057. 10134,10252. 10354,10598. 10705,10790. 10885,  
 10984. 11087,11464. 11584,12093. 12217,12298. 12388,  
 12540. 12596,12672. 12769,12869. 12920,13040. 13170,  
 13200. 13443,13818. 13961,14057. 14149,14241. 14432,  
 14513. 14581,14676. 14755,14918. 15026,15369. 15440,  
 15616. 15696,15772. 15892,15965. 16088,16167. 16372,  
 16450. 17510)  
 /genes="c316G12.1"  
 /product="c316G12.1 (KIAA0734 (C2 domain protein))"  
 /notes="match: cDNAs: Em:AB018277 Em:AF038191  
 match: ESTs: Em:T77480 Em:HI8551 Em:F13300 Em:R61025  
 Em:R24790 Em:R85179 Em:AT217078 Em:AA078749 Em:AI696810  
 Em:AI086907 Em:HI15802 Em:AA293357 Em:R46334 Em:R81844  
 Em:AI301780 Em:R45522 Em:AA782001 Em:R60906 Em:AA716721  
 Em:AI613004 Em:AA716524 Em:C00805 Em:AI111375 Em:AA481839  
 Em:W01878 Em:AI146703 Em:AA627545 Em:AA079823 Em:AI240952  
 Em:AA258690 Em:AA482458 Em:AA682952 Em:AI638343  
 Em:AI693236 Em:AI537389 Em:HI8444 Em:AI358964 Em:AI564977  
 Em:AA302073"  
 /evidence=not experimental

##### CDS

join(6621. 6751,6973. 7060,7235. 7315,7567. 7674,  
 8993. 9052,9200. 9266,9365. 9560,9833. 9937,10057. 10134,  
 10252. 10354,10598. 10705,10790. 10885,10984. 11087,  
 11464. 11584,12093. 12217,12298. 12388,12540. 12596,  
 12672. 12769,12869. 12920,13040. 13170,13320. 13443,  
 13818. 13961,14057. 14149,14241. 14432,14513. 14581,  
 14676. 14755,14918. 15026,15369. 15440,15616. 15696,  
 15772. 15892,15965. 16088,16167. 16372,16450. 16553)  
 /genes="c316G12.1"  
 /codon\_start=1  
 /evidence=not experimental

/product="c316G12.1 (KIAA0734 (C2 domain protein))"  
 /protein\_id="CAB56182.1"  
 /db\_xref="GI:5912546"  
 /db\_xref="SPTREMBL:O94839"  
 /translation="MSTLIDIKSSVLRVQVQVCFSPRRTEQDPSGASADPQEPATGAW  
 KPGDGVFFAHMLMLKKGQGFCEVPLRSGSPAPPEVDFSLGRLALAPEVE  
 MYEALYVLYRAGTMGPDQVDDEALLSYLQVFGTSLEAHTAIEVRKAKPT  
 ALKVSVWRKALLILKPNQFDFYCNMLGILPASDATREPKQRFKRGSKRGTP  
 LPAKCTQVTEVKSSTLNPWYKEHFLFEIDVSTDLHLDIWDHDDVSLVACRKLNE  
 VIGLKGMGYFQXQIVKASARANGTAGTBTEDTDLGLCLNI PVREVPVAGVDRWPKLEP

##### misc\_feature

7785. 8651  
 /genes="c316G12.1"  
 /notes="Tandem repeat. Forced join. Region of tandem  
 repeat, typical sequence of each element:  
 GGTGCACACACACACACAGGTGCA Restriction digest data  
 (BAMHI) suggest approximately 100 base pairs are missing  
 from this assembly."  
 7788. 8686  
 /notes="29 copies 31 mer 65% conserved"  
 7817. 8048  
 /notes="8 copies 29 mer 72% conserved"  
 7819. 7953  
 /notes="5 copies 27 mer 70% conserved"  
 7937. 8299  
 /notes="11 copies 33 mer 66% conserved"  
 7968. 8270  
 /notes="3 copies 101 mer 72% conserved"  
 8224. 8397  
 /notes="6 copies 29 mer 70% conserved"  
 8417. 8474  
 /notes="2 copies 29 mer 98% conserved"  
 8450. 8690  
 /notes="7 copies 33 mer 68% conserved"  
 8478. 8635  
 /notes="2 copies 29 mer 93% conserved"  
 8538. 8673  
 /notes="4 copies 29 mer 75% conserved"  
 8592. 8617  
 /notes="13 copies 2 mer ca 92% conserved"  
 8596. 8676  
 /notes="3 copies 27 mer 84% conserved"  
 8637. 8700  
 /notes="22 copies 2 mer ac 75% conserved"  
 11087. 11288  
 /notes="101 copies 2 mer 99 56% conserved"  
 11215. 11313  
 /notes="3 copies 33 mer 77% conserved"  
 14773. 14974  
 /notes="2 copies 101 mer 93% conserved"  
 15751. 16315  
 /genes="c316G12.1"  
 /notes="CpG island"  
 /evidence=not experimental

##### misc\_feature

15751. 16315

/genes="c316G12.1"

/notes="CpG island"

/evidence=not experimental

/notes="match: STS: Em:R60906"

complement(join(17311. 17679,17960. 18023,18129. 18305,

18878. 19071,19450. 19669,19764. 19875))

/genes="c316G12.2"

complement(join(17311. 17679,17960. 18023,18129. 18305,

18878. 19071,19450. 19669,19764. 19875))

/genes="c316G12.3"

/product="c316G12.2 (novel protein similar to predicted

yeast worm and archae-bacterial proteins)"

/notes="match: cDNAs: Em:AI109401

match: ESTs: Em:AI375098 Em:AI414848 Em:AI660316

Em:AI143673 Em:AI435931 Em:AI693236 Em:AI091298

Em:AI675241 Em:AA215709 Em:AI524203 Em:AI684711

Em:AI571317 Em:AI094880 Em:AA317548 Em:AI341540 Em:H98014

```

Em:AA716721 Em:AI659324 Em:AI224948 Em:AA410348
Em:AI086907 Em:AI304322 Em:AI146703 Em:W01878 Em:AI111375
Em:AA262969 Em:C00805 Em:H46334 Em:D81844 Em:AI696810
Em:AA716524 Em:AI217078 Em:AA079823 Em:AA627545
Em:AA483839 Em:AI560264 Em:AA258690 Em:AI240952
Em:AA484458 Em:HI5802 Em:AA302073 Em:AI423636 Em:R45522
Em:AA293357 Em:AI638343 Em:AI301780 Em:R60906 Em:AI672681
Em:AI613004 Em:AA782001 Em:F31747 Em:AI537389 Em:HI8444
Em:AI358964 Em:AI564977
/evidence=not_experimental

Query Match      87.0%; Score 17.4; DB 9; Length 37476;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGGCCCCCAGCCACCA 19
Db 6760 CGAGGGCCCCCAGCCACCA 6778
|||||
|||||

RESULT 12
AF003626/c
LOCUS
DEFINITION
Homo sapiens chromosome X clone ICRFXc104-M0525, LLOXNC01-C1233,
ICRFXC104-C1284, Qc-3C1, LLOXNC01-B1439, LLOXNC01-220B3,
LLOXNC01-57B6, Qc-12C11 map q28, complete sequence.
ACCESSION
AF003626
VERSION
AF003626.2 GI:27476107
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 86834)
Bleichschmidt, K., Rosenthal, A. and Platzer, M.
Chromosome X genomic sequence
Unpublished
2 (bases 1 to 86834)
Bleichschmidt, K., Drescher, B., Nordsiek, G., Schattevoy, R., Knop, A.,
Rosenthal, A., Kioschis, P. and Poustka, A.
Direct Submission
Submitted (09-MAY-1997) Genome Analysis, Institute for Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
3 (bases 1 to 86834)
Platzer, M.
Direct Submission
Submitted (03-JAN-2003) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
On Jan 3, 2003 this sequence version replaced gi:2121302.
-----
Center: Institute of Molecular Biotechnology
Center code: IVB
Web site: http://genome.imb-jena.de/
Contact: gscj-submit@genome.imb-jena.de
-----
Project Information
Center project name: x99+
Center clone name: ICRFXc104-M0525 to Qc-12C11
-----
Summary Statistics
Sequencing vector: M13, pUC18; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 86833 bases at least Q40
Consensus quality: 86833 bases at least Q30
Consensus quality: 86834 bases at least Q20
Quality coverage: 24.18%
-----
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest.

```

-----  
Neighboring sequence information:  
This entry is overlapped by RP11-316K19, RP4-687A5, LLOXNC01-223D9  
and RP4-741010.  
-----

#### Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

FEATURES  
Source

#### Location/Qualifiers

1..86834  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="q28"  
/clone="ICRFXC104-M0525, LLOXNC01-C1233, ICRFXc104-C1284, Qc-3C1, LLOXNC01-B1439, LLOXNC01-220B3, LLOXNC01-57B6, Qc-12C11"  
1..11008  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="LLOXNC01-223D"  
/clone\_lib="Lawrence Livermore human X cosmid library"  
/notes="overlapping clone"  
1..22000  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RP11-316K1"  
/clone\_lib="RP4-687A"  
/notes="overlapping clone"  
1..26252  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RP4-687A"  
/clone\_lib="RP4-687A"  
/notes="overlapping clone"  
1..45419  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="ICRFXC104-M0525"  
/clone\_lib="ICRF human X specific cosmid library no.104"  
2178..45419  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="LLOXNC01-C1233"  
/clone\_lib="Lawrence Livermore human X cosmid library"  
10530..53681  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="ICRFXC104-C1284"  
/clone\_lib="ICRF human X specific cosmid library no.104"  
11279..47432  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="Qc-3C1"  
/clone\_lib="QIZ-derived Xq27.3-qter cosmid library"  
14100..59047  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"

```

/db xref="taxon:9606"
/clone="LLOXNC01-B1439"
/clone lib="Lawrence Livermore human X cosmid library"
43701..80327
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db xref="taxon:9606"
/clone="LLOXNC01-220B3"
/clone lib="Lawrence Livermore human X cosmid library"
49823..86834
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db xref="taxon:9606"
/clone="Qc-12C11"
/clone lib="Q12-derived Xq27.3-Xqter cosmid library"
52548..86834
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db xref="taxon:9606"
/clone="RP4-74101"
/clone lib="RPCI human PAC 4"
/notes="overlapping clone"
74537..83583
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db xref="taxon:9606"
/clone="LLOXNC01-57B6"
/clone lib="Lawrence Livermore human X cosmid library"
1029..1048
/notes="deleted in clone: RP4-687A5, deleted in clone:
LLOXNC01-223D9, deleted in clone: ICRFXC104-M0525"
/replace="G"
1537
/notes="A substituted in clone: RP11-316K19"
/replace="G"
3051
/notes="T substituted in clone: RP11-316K19"
/replace="G"
3156
/notes="C substituted in clone: RP11-316K19"
/replace="A"
4270..4273
/notes="deleted in clone: RP4-687A5, deleted in clone:
LLOXNC01-223D9, deleted in clone: LLOXNC01-C1233"
/replace="G"
4299..4306
/notes="deleted in clone: ICRFXC104-M0525"
/replace="G"
4302..4303
/notes="deleted in clone: RP4-687A5, deleted in clone:
LLOXNC01-C1233, deleted in clone: LLOXNC01-223D9"
/replace="G"
4336..4361
/notes="deleted in clone: RP4-687A5"
/replace="G"
4350..4361
/notes="deleted in clone: RP11-316K19"
/replace="G"
4378
/notes="deleted in clone: LLOXNC01-223D9, deleted in
clone: RP4-687A5"
/replace="G"
9658
/notes="T substituted in clone: LLOXNC01-C1233"
/replace="C"
11514
/notes="deleted in clone: LLOXNC01-C1233, deleted in
clone: ICRFXC104-C1284, deleted in clone:
ICRFXC104-M0525"
/replace="G"
12953
/notes="A substituted in clone: RP11-316K19"
/replace="G"

```

```

variation 13065
/notes="G substituted in clone: RP11-316K19"
/replace="T"

Query Match 87.0%; Score 17.4; DB 9; Length 86834;
Best Local Similarity 94.7%; Pred. No. 1.le+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCGAGCCGCCACA 19
Db 16859 CGAGGTGCCAGCCGCCACA 16841

RESULT 13
AL354677/c 97834 bp DNA linear HTG 13-JUN-2001
LOCUS Homo sapiens chromosome X clone RP11-479P15, 27 unordered pieces.
DEFINITION AL354677
ACCESSION AL354677
VERSION AL354677.3 GI:9931730
HTG: HTGS PHASE1: HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Melay, K.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 28, 2000 this sequence version replaced gi:9963680.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA479P15
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 78464 bases at least Q40
Consensus quality: 85960 bases at least Q30
Consensus quality: 90075 bases at least Q20
Insert size: 95234; sum-of-contigs
Insert size: 197762; agarose-fp
Quality coverage: 1.99X in Q20 bases; sum-of-contigs Quality
Coverage: 1.47X in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 3673: contig of 3673 bp in length
* 3674 3773: gap of 100 bp
* 3774 6066: contig of 2293 bp in length
* 6067 6166: gap of 100 bp
* 6167 9708: contig of 3540 bp in length
* 9707 9808: gap of 100 bp
* 9807 21235: contig of 11429 bp in length
* 21236 21336: gap of 100 bp
* 21336 24749: contig of 3414 bp in length
* 24750 24850: gap of 100 bp
* 24850 27355: contig of 2506 bp in length
* 27356 27455: gap of 100 bp
* 27456 31116: contig of 3661 bp in length
* 31117 31216: gap of 100 bp
* 31217 35127: contig of 3911 bp in length

```

```

* 35128 35227: gap of 100 bp
* 35228 38448: contig of 3221 bp in length
* 38449 38548: gap of 100 bp
* 38549 40769: contig of 2221 bp in length
* 40770 40869: gap of 100 bp
* 40870 45236: contig of 4367 bp in length
* 45237 45336: gap of 100 bp
* 45337 47792: contig of 2436 bp in length
* 47793 47892: gap of 100 bp
* 47893 53266: contig of 5374 bp in length
* 53267 53366: gap of 100 bp
* 53367 58271: contig of 4905 bp in length
* 58272 58371: gap of 100 bp
* 58372 63797: contig of 5426 bp in length
* 63798 63897: gap of 100 bp
* 63898 66275: contig of 2378 bp in length
* 66276 66375: gap of 100 bp
* 66376 68543: contig of 2168 bp in length
* 68544 70687: contig of 2044 bp in length
* 70688 70787: gap of 100 bp
* 70788 73405: contig of 2618 bp in length
* 73406 73505: gap of 100 bp
* 73506 76943: contig of 3438 bp in length
* 76944 77043: gap of 100 bp
* 77044 79123: contig of 2080 bp in length
* 79124 79223: gap of 100 bp
* 79224 83079: contig of 3856 bp in length
* 83080 83179: gap of 100 bp
* 83180 85477: contig of 2298 bp in length
* 85478 85577: gap of 100 bp
* 85578 88929: contig of 3352 bp in length
* 88930 89029: gap of 100 bp
* 89030 91739: contig of 2710 bp in length
* 91740 91839: gap of 100 bp
* 91840 94358: contig of 2419 bp in length
* 94359 94358: gap of 100 bp
* 94359 97834: contig of 3476 bp in length.
* 94359 97834: contig of 3476 bp in length.
FEATURES
    source
        1..97834
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="X"
            /clone="RP11-479P15"
            /clone_lib="RPC1-11.2"
            1..3673
                /note="assembly fragment:00141"
                fragment_chain:1
                3774..5066
                    /note="assembly fragment:00636"
                    fragment_chain:1
                    6167..9706
                        /note="assembly fragment:00363"
                        fragment_chain:2
                        9807..21235
                            /note="assembly fragment:00564"
                            fragment_chain:2
                            21336..24749
                                /note="assembly fragment:00449"
                                fragment_chain:3
                                24850..27355
                                    /note="assembly fragment:00808"
                                    fragment_chain:3
                                    27456..31116
                                        /note="assembly fragment:00076"
                                        31217..35127
                                            /note="assembly fragment:00114"
                                            35228..38448
                                                /note="assembly fragment:00143"
                                                38549..40769
                                                    /note="assembly fragment:00173"
                                                    40870..45236

```

```

misc_feature /note="assembly_fragment:00176"
45337..47792
misc_feature /note="assembly_fragment:00225"
47893..53266
misc_feature /note="assembly_fragment:00265"
53367..58271
misc_feature /note="assembly_fragment:00355"
58372..63797
misc_feature /note="assembly_fragment:00400"
63898..66275
misc_feature /note="assembly_fragment:00423"
66376..68543
misc_feature /note="assembly_fragment:00460"
68544..70687
misc_feature /note="assembly_fragment:00461"
70788..73405
misc_feature /note="assembly_fragment:00464"
73506..76943
misc_feature /note="assembly_fragment:00536"
77044..79123
misc_feature /note="assembly_fragment:00586"
79224..83079
misc_feature /note="assembly_fragment:00631"
83180..85477
misc_feature /note="assembly_fragment:00707"
85578..88929
misc_feature /note="assembly_fragment:00763"
89030..91739
misc_feature /note="assembly_fragment:00773"
91840..94258
misc_feature /note="assembly_fragment:00830"
94359..97834
misc_feature /note="assembly_fragment:00894"
97834..97834
ORIGIN
Query Match 87.0%; Score 17.4; DB 2; Length 97834;
Best Local Similarity 94.7%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CGAGGCGCCCGCCGCCACA 19
Db 57840 CGAGGCGCCCGCCGCCACA 57822

```

```

RESULT 14
AC092269 105589 bp DNA linear HTG 04-SEP-2001
LOCUS Homo sapiens chromosome 5 clone CTB-78L15, WORKING DRAFT SEQUENCE.
DEFINITION AC092269
AC092269
VERSION AC092269.2 GI:15290438
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 105589)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 105589)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL DOE Joint Genome Institute.
COMMENT Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 25, 2001 this sequence version replaced gi:14589458.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 94435
Center clone name: CIT978SKB_78L15

```

```

-----
Summary Statistics
Consensus quality: 104529 bases at least Q40
Consensus quality: 105329 bases at least Q30
Consensus quality: 105517 bases at least Q20
Estimated insert size: 107000; agarose-fp estimation
Estimated insert size: 105589; sum-of-contigs estimation
Quality coverage: 9.12 in Q20 bases; agarose-fp estimation
Quality coverage: 9.24 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 105589: contig of 105589 bp in length.
FEATURES             Location/Qualifiers
     source            1. .105589
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="5"
                        /clone="CTB-78L15"
                        /clone_lib="CalTech human BAC library B"

ORIGIN
Query Match      87.0%; Score 17.4; DB 2; Length 105589;
Best Local Similarity 94.7%; Pred. NO. 1.le+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GAGGCGCCAGCCCAACAA 20
Db 38624 GAGGCGCCAGCCCAACAA 38642

RESULT 15
AF187000/c
LOCUS             129679 bp      DNA      linear      PRI 03-SEP-2002
DEFINITION       Homo sapiens chromosome 8 clone RPI-316L14 map 8q24.2, complete
sequence.
ACCESSION        AF187000
VERSION          AF187000.4  GI:22657510
KEYWORDS         HTG.
SOURCE           Homo sapiens (human)
ORGANISM         Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 129679)
Wen, G., Baumgart, C., Blechschmidt, K., Dette, M., Jahn, N.,
Lehmann, R., Menzel, U., Polley, A., Reichwald, K., Schillabel, M.B.,
Schudy, A., Siddiqui, R., Taudien, S., Schlegelberger, B., Siebert, R.,
Rosenthal, A. and Platzzer, M.
Unpublished
2 (bases 1 to 129679)
Reichwald, K., Schlegelberger, B., Siebert, R., Baumgart, C.,
Menzel, U., Jahn, N. and Rosenthal, A.
Direct Submission
Submitted (10-SEP-1999) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
3 (bases 1 to 129679)
Genome Sequencing Center Jena.
Direct Submission
Submitted (24-MAY-2000) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
4 (bases 1 to 129679)
Wen, G. and Platzzer, M.
Direct Submission
Submitted (05-FEB-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
5 (bases 1 to 129679)
Lagemann, D. and Platzzer, M.
-----
Direct Submission
Submitted (03-SEP-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
On Sep 3, 2002 this sequence version replaced GI:18497321.
-----
Genome Center
Center: Institute of Molecular Biotechnology
Center code: IMB
Web site: http://genome.imb-jena.de/
Contact: gscj-submit@genome.imb-jena.de
-----
Project Information
Center project name: H244
Center clone name: RPI-316L14
-----
Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 127436 bases at least Q40
Consensus quality: 128952 bases at least Q30
Consensus quality: 129513 bases at least Q20
Quality coverage: 7.27x
-----
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
Location/Qualifiers
     source            1. .129679
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="8"
                        /map="8q24.2, complete sequence."
                        /clones="RPI-316L14"
3571
misc_feature       /note="low quality region , RPI-316L14"
7007
misc_feature       /note="low quality region , RPI-316L14"
7027..7428
misc_feature       /note="single stranded/single chemistry region"
7429..7652
misc_feature       /note="single stranded/single chemistry region"
7663..7968
misc_feature       /note="single stranded/single chemistry region"
10725
misc_feature       /note="low quality region , RPI-316L14"
10813
misc_feature       /note="low quality region , RPI-316L14"
14411..14473
misc_feature       /note="single stranded/single chemistry region"
16701..16991
misc_feature       /note="single stranded/single chemistry region"
16756
misc_feature       /note="low quality region , RPI-316L14"
16812
misc_feature       /note="low quality region , RPI-316L14"
16814
misc_feature       /note="low quality region , RPI-316L14"
16894
misc_feature       /note="low quality region , RPI-316L14"

```



```

misc_feature 15897 /note="low quality region , RP1-316L14"
misc_feature 15898 /note="low quality region , RP1-316L14"
misc_feature 22591. .22776 /note="single stranded/single chemistry region"
misc_feature 28909 /note="low quality region , RP1-316L14"
misc_feature 28922 /note="low quality region , RP1-316L14"
misc_feature 28968 /note="low quality region , RP1-316L14"
misc_feature 29023. .29024 /note="low quality region , RP1-316L14"
misc_feature 29052 /note="low quality region , RP1-316L14"
misc_feature 29055. .29059 /note="low quality region , RP1-316L14"
misc_feature 29070. .29071 /note="low quality region , RP1-316L14"
misc_feature 29127 /note="low quality region , RP1-316L14"
misc_feature 29129 /note="low quality region , RP1-316L14"
misc_feature 29131 /note="low quality region , RP1-316L14"
misc_feature 29135 /note="low quality region , RP1-316L14"
misc_feature 29137 /note="low quality region , RP1-316L14"
misc_feature 29140 /note="low quality region , RP1-316L14"
misc_feature 29144 /note="low quality region , RP1-316L14"
misc_feature 29147 /note="low quality region , RP1-316L14"
misc_feature 29153. .29154 /note="low quality region , RP1-316L14"
misc_feature 29160 /note="low quality region , RP1-316L14"
misc_feature 30845 /note="low quality region , RP1-316L14"
misc_feature 30882. .31175 /note="single stranded/single chemistry region"
misc_feature 31960. .32018 /note="single stranded/single chemistry region"
misc_feature 31962 /note="single stranded/single chemistry region"
misc_feature 31964 /note="low quality region , RP1-316L14"
misc_feature 31967. .31969 /note="low quality region , RP1-316L14"
misc_feature 31971. .31972 /note="low quality region , RP1-316L14"
misc_feature 31980 /note="low quality region , RP1-316L14"
misc_feature 31986. .31991 /note="low quality region , RP1-316L14"
misc_feature 31993. .31996 /note="low quality region , RP1-316L14"
misc_feature 31999. .32001 /note="low quality region , RP1-316L14"
misc_feature 32014. .32016 /note="low quality region , RP1-316L14"
misc_feature 32018 /note="low quality region , RP1-316L14"
misc_feature 33606. .33659 /note="single stranded/single chemistry region"
misc_feature 35711. .35954 /note="single stranded/single chemistry region"
misc_feature 40328. .40352 /note="pcr product sequence only , RP1-316L14"
misc_feature 40409. .40428

```

```

misc_feature /note="single clone coverage"
40429. .40504 /note="single stranded/single chemistry region"
40443 /note="low quality region , RP1-316L14"
40459. .40460 /note="low quality region , RP1-316L14"
40491 /note="low quality region , RP1-316L14"
43087 /note="low quality region , RP1-316L14"
43185 /note="low quality region , RP1-316L14"
43198. .43304 /note="single stranded/single chemistry region"
43211. .43212 /note="low quality region , RP1-316L14"
43301 /note="low quality region , RP1-316L14"
43851 /note="low quality region , RP1-316L14"

Query Match 87.0%; Score 17.4; DB 9; Length 129679;
Best Local Similarity 94.7%; Pred.No.1.le+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GAGCGCCGACGCCACAA 20
Db 57072 GAGCGCCGACGCCACAA 57054
|||||
|||||

RESULT 16
AC069493
LOCUS
DEFINITION Homo sapiens chromosome 2 clone RP11-591H15 map 2, WORKING DRAFT
SEQUENCE, 3 unordered pieces.
AC069493
AC069493.3 GI:15291097
VERSION HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 154807)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepey,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Menius,L., Mihova,T., Miranda,C., Mienna,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-JUN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 26, 2001 this sequence version replaced gi:8576270.
All repeats were identified using RepeatMasker.

```

Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L9126  
Center clone name: 591.H.15  
----- Summary Statistics  
Sequencing vector: M13; M77815; 37% of reads  
Sequencing vector: Plasmid; n/a; 63% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 154136 bases at least Q40  
Consensus quality: 154423 bases at least Q30  
Consensus quality: 154547 bases at least Q20  
Insert size: 145000; agarose-fp  
Insert size: 154607; sum-of-contigs  
Quality coverage: 11.9 in Q20 bases; agarose-fp  
Quality coverage: 11.2 in Q20.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\*  
\* 1 5367: contig of 5367 bp in length  
\* 5368 5467: gap of 100 bp  
\* 5468 60792: contig of 55325 bp in length  
\* 60793 60892: gap of 100 bp  
\* 60893 154807: contig of 93915 bp in length.  
\*  
\* Location/Qualifiers  
FEATURES  
source  
1. 154807  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-591H15"  
/clone\_lib="RP11-11 Human Male BAC"  
misc\_feature  
1. 5367  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
misc\_feature  
5468..60792  
/note="assembly\_fragment"  
misc\_feature  
60893..154807  
/note="assembly\_fragment  
clone\_end:T7  
vector\_side:right"  
ORIGIN  
Query Match 87.0%; Score 17.4; DB 2; Length 154807;  
Best Local Similarity 94.7%; Pred. No. 1.le+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 GAGCGCGCCAGCCCCACAA 20  
Db 153649 GAGCGCGCCAGCCCCACAA 153667  
RESULT 17  
AC008393/c  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone CTC-241N9, complete sequence.  
ACCESSION AC008393  
VERSION AC008393.7 GI:18921273  
KEYWORDS HTG.  
SOURCE  
Homo sapiens (human)

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 166847)  
AUTHORS  
DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE  
Direct Submission  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 166847)  
AUTHORS  
DOE Joint Genome Institute.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE  
3 (bases 1 to 166847)  
AUTHORS  
DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (26-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
COMMENT  
On Feb 25, 2002 this sequence version replaced gi:15290282.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.6% of Sequence;  
Estimated Total Number of Errors is 0.5.  
NOTE: This insert is not the entire sequence of the clone (entire  
sequence is 178kb). It is clipped at the overlap with AC010285. The  
number of bases overlapped is 24225.  
FEATURES  
Location/Qualifiers  
1. 166847  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTC-241N9"  
ORIGIN  
Query Match 87.0%; Score 17.4; DB 9; Length 166847;  
Best Local Similarity 94.7%; Pred. No. 1.le+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 GAGCGCGCCAGCCCCACAA 20  
Db 81921 GAGCGCGCCAGCCCCACAA 81903  
RESULT 18  
AC023924  
LOCUS  
DEFINITION Homo sapiens chromosome 4 clone RP11-462J14 map 4, \*\*\* SEQUENCING  
IN PROGRESS \*\*\*, 42 unordered pieces.  
ACCESSION AC023924  
VERSION AC023924.2 GI:7331424  
KEYWORDS HTG: HTGS PHASE1.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 175511)  
AUTHORS  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE  
Homo sapiens chromosome 4, clone RP11-462J14  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 175511)  
AUTHORS  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavsky, L.,  
Bouckhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,  
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
DeArillano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,  
Fenster, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kuan, L., Karatas, A.,  
Klein, J., Landers, T., Largocque, K., Lehocsky, J., Levine, R.,  
Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,

McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J.,  
Meneus, L., Mikova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,  
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T.M.,  
Peterson, K., Pierre, N., Pisci, C., Pollara, V., Raymond, C.,  
Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,  
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,  
Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,  
Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and  
Zody, M.

Direct Submission  
Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 28, 2000 this sequence version replaced gi:7009813.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Project name: L4974  
Center project name: L4974  
Center clone name: 462\_J\_14  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 42 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1071: contig of 1071 bp in length  
\* 1072 1171: gap of 100 bp  
\* 1172 2441: contig of 1270 bp in length  
\* 2442 2541: gap of 100 bp  
\* 2542 3921: contig of 1379 bp in length  
\* 3921 4020: gap of 100 bp  
\* 4021 5112: contig of 1092 bp in length  
\* 5113 5213: gap of 100 bp  
\* 5213 6857: contig of 1644 bp in length  
\* 6857 8297: gap of 100 bp  
\* 8297 8330: contig of 1373 bp in length  
\* 8330 8429: gap of 100 bp  
\* 8430 10443: contig of 2014 bp in length  
\* 10444 10543: gap of 100 bp  
\* 10544 11871: contig of 1328 bp in length  
\* 11872 11872: gap of 100 bp  
\* 11872 12442: contig of 271 bp in length  
\* 12443 12342: gap of 100 bp  
\* 12343 14098: contig of 1756 bp in length  
\* 14099 14198: gap of 100 bp  
\* 14199 15712: contig of 1514 bp in length  
\* 15713 15813: gap of 100 bp  
\* 15813 18266: contig of 2454 bp in length  
\* 18267 18366: gap of 100 bp  
\* 18367 20390: contig of 1924 bp in length  
\* 20391 20391: gap of 100 bp  
\* 20391 22569: contig of 2179 bp in length  
\* 22570 22669: gap of 100 bp  
\* 22670 25385: contig of 2716 bp in length  
\* 25386 25485: gap of 100 bp  
\* 25486 28984: contig of 3399 bp in length  
\* 28985 31777: contig of 2793 bp in length  
\* 31778 31877: gap of 100 bp  
\* 31878 34560: contig of 2683 bp in length  
\* 34561 34661: gap of 100 bp  
\* 34661 37141: contig of 2481 bp in length  
\* 37142 37241: gap of 100 bp  
\* 37242 40350: contig of 3109 bp in length

\* 40351 40450: gap of 100 bp  
\* 40451 43420: contig of 2970 bp in length  
\* 43421 43520: gap of 100 bp  
\* 43521 46331: contig of 2811 bp in length  
\* 46332 46431: gap of 100 bp  
\* 46432 49168: contig of 2737 bp in length  
\* 49169 49268: gap of 100 bp  
\* 49269 52366: contig of 3098 bp in length  
\* 52367 52466: gap of 100 bp  
\* 52467 56370: gap of 100 bp  
\* 56371 59259: contig of 2889 bp in length  
\* 59260 63099: contig of 3740 bp in length  
\* 63100 63199: gap of 100 bp  
\* 63200 67884: contig of 4685 bp in length  
\* 67885 71879: contig of 3895 bp in length  
\* 71880 71979: gap of 100 bp  
\* 71980 77510: contig of 5531 bp in length  
\* 77511 81594: contig of 3984 bp in length  
\* 81595 87445: contig of 5751 bp in length  
\* 87446 92288: contig of 4743 bp in length  
\* 92289 92388: gap of 100 bp  
\* 92389 97883: contig of 5495 bp in length  
\* 97884 104615: contig of 6632 bp in length  
\* 104616 104716: gap of 100 bp  
\* 104717 112913: contig of 8198 bp in length  
\* 112914 113013: gap of 100 bp  
\* 113014 121166: contig of 8153 bp in length  
\* 121167 121266: gap of 100 bp  
\* 121267 129653: contig of 8387 bp in length  
\* 129654 139142: contig of 9389 bp in length  
\* 139143 139242: gap of 100 bp  
\* 139243 148958: contig of 9716 bp in length  
\* 148959 149058: gap of 100 bp  
\* 149059 160677: contig of 11618 bp in length  
\* 160677 160777: gap of 100 bp  
\* 160777 175511: contig of 14735 bp in length.

FEATURES  
Location/Qualifiers

1. .175511  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-462J14"  
/clone\_lib="RPC1-11 Human Male BAC"  
misc\_feature  
1. .1071  
/note="assembly\_fragment"  
misc\_feature  
1172. .2441  
/note="assembly\_fragment"  
misc\_feature  
2542. .3920  
/note="assembly\_fragment"  
misc\_feature  
4021. .5112  
/note="assembly\_fragment"  
misc\_feature  
5213. .6856  
/note="assembly\_fragment"  
misc\_feature  
6957. .8329  
/note="assembly\_fragment"  
misc\_feature  
8430. .10443  
/note="assembly\_fragment"  
misc\_feature  
10544. .11871  
/note="assembly\_fragment"  
misc\_feature  
11972. .12242  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right

Mon Jun 21 09:02:04 2004

```

misc_feature      12343..14098
/note="assembly_fragment"
misc_feature      14199..15712
/note="assembly_fragment"
misc_feature      15813..18266
/note="assembly_fragment"
misc_feature      18367..20290
/note="assembly_fragment"
misc_feature      20391..22565
/note="assembly_fragment"
misc_feature      22670..25385
/note="assembly_fragment"
misc_feature      25486..28884
/note="assembly_fragment"
misc_feature      28985..31777
/note="assembly_fragment"
misc_feature      31878..34560
/note="assembly_fragment"
misc_feature      34661..37141
/note="assembly_fragment"
misc_feature      37242..40350
/note="assembly_fragment"
misc_feature      40451..43420
/note="assembly_fragment"
misc_feature      43521..46331
/note="assembly_fragment"
misc_feature      46432..49168
/note="assembly_fragment"

Query Match      87.0%; Score 17.4; DB 2; Length 175511;
Best Local Similarity 94.7%; Pred. No. 1.le+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCAGCCCAACA 20
Db 168379 GAGGCGCCAGCCCAACA 168397

RESULT 19
AC142187
LOCUS      181070 bp DNA linear HTG 24-MAR-2003
DEFINITION Rattus norvegicus clone CH230-368G11, WORKING DRAFT SEQUENCE, 28
unordered pieces.
AC142187
VERSION      GI:29165337
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS      Rattus norvegicus (Norway rat)
SOURCE      Rattus norvegicus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 181070)
Muzny,D,Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Blawalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Cantar,R., Garcia,A., Garner,T., Garza,M.,
Gebrgeorgis,E., Geer,K., Gail,R., Grady,M., Guerra,W., Guvera,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,

```

```

Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensheva,L., Loulsegad,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,N., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., Mcleod,M., McNeill,T., Meenen,E., Milosavljevic,A.,
Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
Morris,S., Munjasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okunolu,G.,
Olarpunsgoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
Perez,A., Perez,L., Pfanckoch,C., Plopper,F., Poindexter,A.,
Popovic,D., Primus,E., Pu,L.-L., Puzo,M., Quiroz,J., Rachlin,E.,
Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
Sisson,I., Sitter,C.D., Smajls,D., Sneed,A., Sodergren,E.,
Song,X.-Z., Sorelle,R., Sosa,J., Steinle,M., Strong,R., Sutton,A.,
Svatek,A., Tabor,P., Taylor,C., Taylor,I., Thomas,N., Thomas,S.,
Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,X.,
Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: KSKU
Center clone name: CH230-368G11
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167171 bases at least Q40
Consensus quality: 170222 bases at least Q30
Consensus quality: 172408 bases at least Q20
Estimated insert size: 172425; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see this://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1309: contig of 1309 bp in length
* 1409: gap of unknown length
* 1410: contig of 1007 bp in length
* 2416: gap of unknown length
* 2417: gap of unknown length
* 2517: contig of 1836 bp in length
* 4352: contig of 1836 bp in length
* 4353: gap of unknown length
* 4452: gap of unknown length
* 4549: contig of 1197 bp in length
* 5650: gap of unknown length
* 5750: contig of 1527 bp in length
* 7276: gap of unknown length
* 7277: gap of unknown length

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```
* 7377 9964: contig of 2588 bp in length
* 9965 10064: gap of unknown length
* 10065 11720: contig of 1656 bp in length
* 11721 11820: gap of unknown length
* 11821 15895: contig of 4075 bp in length
* 15896 15995: gap of unknown length
* 15996 18157: contig of 2162 bp in length
* 18158 18257: gap of unknown length
* 18258 22416: contig of 4159 bp in length
* 22417 22516: gap of unknown length
* 22517 26229: contig of 3713 bp in length
* 26230 26329: gap of unknown length
* 26330 30415: contig of 4086 bp in length
* 30416 30515: gap of unknown length
* 30516 34796: contig of 4231 bp in length
* 34797 34986: gap of unknown length
* 34987 40448: contig of 5552 bp in length
* 40449 40548: gap of unknown length
* 40549 48151: contig of 7603 bp in length
* 48152 48251: gap of unknown length
* 48252 55577: contig of 7326 bp in length
* 55578 55677: gap of unknown length
* 55678 61339: contig of 5862 bp in length
* 61340 61639: gap of unknown length
* 61640 68123: contig of 6384 bp in length
* 68124 75242: contig of 7119 bp in length
* 75243 75342: gap of unknown length
* 75343 80050: contig of 4708 bp in length
* 80051 80150: gap of unknown length
* 80151 91069: contig of 10919 bp in length
* 91070 91169: gap of unknown length
* 91170 99173: contig of 8004 bp in length
* 99174 99273: gap of unknown length
* 99274 106154: contig of 6881 bp in length
* 106155 106254: gap of unknown length
* 106255 117958: contig of 11704 bp in length
* 117959 135029: contig of 16971 bp in length
* 135030 135129: gap of unknown length
* 135130 150364: contig of 15235 bp in length
* 150365 150464: gap of unknown length
* 150465 165222: contig of 14758 bp in length
* 165223 165322: gap of unknown length
* 165323 181070: contig of 15748 bp in length.

FEATURES             Location/Qualifiers
     1..181070
        /organism="Rattus norvegicus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10116"
        /clone="CH230-368G11"

ORIGIN
Query Match      87.0%; Score 17.4; DB 2; Length 181070;
Best Local Similarity 94.7%; Pred. No. 1.1e-03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCAGCCCAACAA 20
Db 150850 GGGGCGCCAGCCCAACAA 150868

RESULT 20
AC124790
LOCUS
DEFINITION Homo sapiens chromosome 8 clone RP11-973F15 map 8, WORKING DRAFT
SEQUENCE, 13 ordered pieces.
ACCESSION AC124790
VERSION AC124790.1 GI:21431211
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
```

```
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 182765)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Bouckgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Farco,S., Ferreira,P., Fitzgerald,M., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Govett,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Illiev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kellis,C., Laroque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
O'Neil,D., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schupbach,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2365
Center clone name: 973F15
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 178183 bases at least Q40
Consensus quality: 180152 bases at least Q30
Consensus quality: 180903 bases at least Q20
Insert size: 194000; agarose-fp
Quality coverage: 6.2 in Q20 bases; agarose-fp
Quality coverage: 6.7 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 14137: contig of 14137 bp in length
* 14138 14237: gap of 100 bp
* 14238 15333: contig of 1096 bp in length
* 15334 15433: gap of 100 bp
* 15434 16426: contig of 993 bp in length
* 16427 16526: gap of 100 bp
* 16527 17835: contig of 1309 bp in length
* 17836 17935: gap of 100 bp
* 17936 67355: contig of 49420 bp in length
* 67356 67455: gap of 100 bp
```

```

* 67456 69673: contig of 2218 bp in length
* 69674 69773: gap of 100 bp
* 69774 72978: contig of 3205 bp in length
* 72979 73078: gap of 100 bp
* 73079 82532: contig of 9454 bp in length
* 82533 82632: gap of 100 bp
* 82633 95585: contig of 12953 bp in length
* 95586 95686: gap of 100 bp
* 95687 117061: contig of 21376 bp in length
* 117062 117162: gap of 100 bp
* 117163 149013: contig of 31852 bp in length
* 149014 149113: gap of 100 bp
* 149114 181372: contig of 32259 bp in length
* 181373 181472: gap of 100 bp
* 181473 182765: contig of 1293 bp in length.

```

Location/Qualifiers

```

1. .182765
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="8"
/maps="8"
/clone="RP11-973F15"
/clone_lib="RP11-973F15 Human Male BAC"
1. .14137
/clone="assembly_fragment"
/clone_end="SP6"
vector_side="left"
14238..15333
/note="assembly_fragment"
15434..16426
/note="assembly_fragment"
16527..17835
/note="assembly_fragment"
17936..67355
/note="assembly_fragment"
67456..69673
/note="assembly_fragment"
69774..72978
/note="assembly_fragment"
73079..82532
/note="assembly_fragment"
82633..95585
/note="assembly_fragment"
95686..117061
/note="assembly_fragment"
117162..149013
/note="assembly_fragment"
149114..181372
/note="assembly_fragment"
181473..182765
/note="assembly_fragment"
-clone_end="T7"
vector_side="right"

```

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

# SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 186908)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 8, clone RP11-973F15

Unpublished

2 (bases 1 to 186908)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Bloom, T., Bonuskavkiy, L.,

Boukhatat, B., Brown, A., Camarato, J., Campopiano, A., Chang, J.,

Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,

Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Faro, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D.,

Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,

Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., LaRocque, K.,

Lamarez, R., Landers, T., Lenoczky, J., Levine, R., Lindblad-Toh, K.,

Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,

Mathews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,

Menues, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,

Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,

O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,

Pollara, V., Raymond, C., Retia, R., Rieback, M., Riley, R., Rise, C.,

Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,

Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,

Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,

Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J.,

Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,

Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIER

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L27366

Center clone name: 973 F.15

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 178955 bases at least Q40

Consensus quality: 182683 bases at least Q30

Consensus quality: 184121 bases at least Q20

Insert size: 194000; agarose-fp

Insert size: 184808; sum-of-contigs

Quality coverage: 5.8 in Q20 bases; agarose-fp

Quality coverage: 6.1 in Q20 bases; sum-of-contigs

# ORIGIN

Query Match

Best Local Similarity

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Score 17.4; DB 2; Length 182765;

2 GAGCGCCGACGCCCCACAA 20

|||||

50578 GAGCGCGACGCCCCACAA 50596

|||||

AC124791

LOCUS

DEFINITION

SEQUENCE, 22 ordered pieces.

AC124791

AC124791.1 GI:21431212

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

KEYWORDS

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 186908)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 8, clone RP11-973F15

Unpublished

2 (bases 1 to 186908)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Bloom, T., Bonuskavkiy, L.,

Boukhatat, B., Brown, A., Camarato, J., Campopiano, A., Chang, J.,

Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,

Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Faro, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D.,

Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,

Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., LaRocque, K.,

Lamarez, R., Landers, T., Lenoczky, J., Levine, R., Lindblad-Toh, K.,

Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,

Mathews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,

Menues, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,

Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,

O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,

Pollara, V., Raymond, C., Retia, R., Rieback, M., Riley, R., Rise, C.,

Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,

Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,

Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,

Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J.,

Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,

Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIER

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L27366

Center clone name: 973 F.15

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 178955 bases at least Q40

Consensus quality: 182683 bases at least Q30

Consensus quality: 184121 bases at least Q20

Insert size: 194000; agarose-fp

Insert size: 184808; sum-of-contigs

Quality coverage: 5.8 in Q20 bases; agarose-fp

Quality coverage: 6.1 in Q20 bases; sum-of-contigs

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 186908)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 8, clone RP11-973F15

Unpublished

2 (bases 1 to 186908)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Bloom, T., Bonuskavkiy, L.,

Boukhatat, B., Brown, A., Camarato, J., Campopiano, A., Chang, J.,

Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,

Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Faro, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D.,

Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,

Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., LaRocque, K.,

Lamarez, R., Landers, T., Lenoczky, J., Levine, R., Lindblad-Toh, K.,

Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,

Mathews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,

Menues, L., Mihova, T., Ml

```

misc_feature      84131..88322
                  /note="assembly_fragment"
88423..94416
                  /note="assembly_fragment"
94517..100512
                  /note="assembly_fragment"
100613..110745
                  /note="assembly_fragment"
110846..129526
                  /note="assembly_fragment"
129627..155933
                  /note="assembly_fragment"
156034..185443
                  /note="assembly_fragment"
185544..186908
                  /note="assembly_fragment"
clone_end:17
vector_side:right

ORIGIN
Query Match      87.0%; Score 17.4; DB 2; Length 186908;
Best Local Similarity 94.7%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2  GAGGGCGCCCGAGCCCAAA 20
          |||||
DB      173029  GAGGGCGACAGCCCAAA 173047

RESULT 22
RME5603647
LOCUS
DEFINITION
RME603647 189333 bp DNA linear BCT 05-JUL-2002
Rhizobium melliloti (Sinorhizobium melliloti) 1021 complete pSymB;
segment 6/5.
ACCESSION
AL603647 AL591985
VERSION
AL603647.1 GI:15141286

```

KEYWORDS  
SOURCE  
Sinorhizobium meliloti (Rhizobium meliloti)

ORGANISM	Sinorhizobium meliloti (Sinorhizobium meliloti)
REFERENCE	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
AUTHORS	Finan, T.M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorholter, F. J., Hernandez-Lucas, I., Becker, A., Cowie, A., Gouzy, J., Golding, B. and Puhler, A.
TITLE	The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endosymbiont Sinorhizobium meliloti
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9889-9894 (2001)
MEDLINE	21396508
PUBMED	11481431
REMARK	epub ahead of print
REFERENCE	2 (bases 1 to 189333)
AUTHORS	Weidner, S.
TITLE	Direct Submission
JOURNAL	Submitted (07-JUN-2001) Weidner S., Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany
COMMENT	Submitted on behalf of Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany and Department of Biology, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4K1 Canada mailto:Stefan.Weidner@Genetik.Uni-Bielefeld.DE pEXO, pSymB.
FEATURES	Location/Qualifiers
source	1. .189333 /organism="Sinorhizobium meliloti" /mol_type="genomic DNA" /strain="1021" /db_xref="taxon:382" /plasmid="pSymB" complement(126. .932) /gene="idnOI OR Smb20692" complement(126. .932)
gene	
CDS	

```

/genes="idnO1 OR SMD20692"
/EC number="1.1.1.69"
/function="Small Molecule Metabolism; Degradation; carbon
compounds"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative 5-Keto-D-gluconate 5-reductase protein"
/protein_id="CAC49799.1"
/db_xref="GI:15141287"
/db_xref="GOA:Q92TU6"
/db_xref="SPTREMBL:Q92TU6"
/translacion="WMGASNAAPRIRGLPALFDLSGRVALITSSGGILGTMALHC
EAGASVILNGREARLAGARAELEHGFVTGSAFDVTKSAEIGRAVADILAEGRID
ILVNRAGIOHPTLPEHFPEDAPRRVLENTLSAFLVGOAVQGMTEREGTILNCSV
QSELARAPTAASKGGLKMLTKGMALDWGRYGRUNGRLARFCYFKTELNSALVSDX
FSTWLEQRTPLGRWGDTGELAAAVFLASASSFVIGHILYVDGGITISCL"
complement(1063..1361)
/genes="SMD20693"
complement(<1063..1361)
/genes="SMD20693"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/notes="Product confidence : hypothetical"
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49800.1"
/db_xref="GI:15141288"
/db_xref="GOA:Q92TU5"
/db_xref="SPTREMBL:Q92TU5"
/translacion="MPDYAASKGVGQAKAFATNGQPCTFRSMELPPGYIATDSTEA
SNDPSRABSIILARIIPRAAGAPRTLPAARRSPWRPHITSRARSSPWTAAAGWA"
1508..1684
/genes="SMD20694"
1508..1684
/genes="SMD20694"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/notes="Product confidence : hypothetical"
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49801.1"
/db_xref="GI:15141289"
/db_xref="GOA:Q92TU4"
/db_xref="SPTREMBL:Q92TU4"
/translacion="WGAGDGFMAALDARMVRVIERLRVADMSAVPDINAGTNPAT
MLGNRCAGFILGTA"
complement(1765..2271)
/genes="SMD20695"
complement(1765..2271)
/genes="SMD20695"
/function="MISCELLANEOUS; Unknown"
/notes="Product confidence : hypothetical"
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49802.1"
/db_xref="GI:15141290"
/db_xref="GOA:Q92TU3"
/db_xref="SPTREMBL:Q92TU3"
/translacion="MTLIPVAGGGLTVAWRILDQAVHAPTWDSEGAAYRVGRWNSKG
VSAVYCSDDPATIAEVAHVKGFRALDVAHMTAAATADAGDVYVVDPKSPVNPNA
RFGIPGAGQQYVGDILLRHRFVAIPSAVSPHSWNILVFLGGAAPAAVAKTQBPFDALD
TRLHPPTG"
complement(2268..2753)

```

```

/genes="SMB20696"
complement(2268..2753)
/genes="SMB20696"
/function="MISCELLANEOUS; Unknown"
/notes="Product confidence : hypothetical"
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49803.1"
/db_xref="GI:15141291"
/db_xref="GOA:Q92TU2"
/db_xref="SPTREMBL:Q92TU2"
/translacion="MALAAKAPPAAGGELQKSIETLGGSRVLAHLTNALDAHELL
LRGLPASALDHLVGNLVFTGKTESLEKAVMSLRTWQRKRTPNKPLSQEQSGRAWKF
AEILVKATDIFGSGAAEQWLERPAGVLQRRPIDLGTTPPGVELVEDHLDRLEYGIY
A"
complement(2855..4549)
/genes="SMB20697"
complement(2855..4549)
/genes="SMB20697"
/function="MACROMOLECULE METABOLISM; Macromolecule
degradation; degradation of proteins, peptides,
glycopeptides"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative peptidase, similar to arginine
utilization protein RecB"
/protein_id="CAC49804.1"
/db_xref="GI:15141292"
/db_xref="GOA:Q92TU1"
/db_xref="SPTREMBL:Q92TU1"
/translacion="MTNAQBEVTSRAGRIDSRVQRFALRMTSPSETGTGCEASFA
DRHLGGLGELPYEPHPODLGLLASHGEPLTRNVVALVRGTGKRTLVMAGHPTDVTST
NYHELKALCDSLADKALIESARTGSEOEERALQDLASGDFLPGRLDMSGL
AVAIACLEQPAADTRQGNLMVLVATPDERESRGNSLRDALPLGVDRFDIAGGIN
LDVTSQGGSEGRAVYAGTIGKLLPFALVIGCSSHAFYFEGVQAQWADALFARLE
GNLADRDNDNDISPPICLEAKLDGDEYITTFERFIANLWYHSMATDALFERFR
EEVLTGANEAIETFAAASAEYKLVGRSAGLPAPKPLLSFOELRAAAARVFGHGFDA
FYAKREFAQSDNPLVATROLTEWLGIARLSGPAIVIGFSGLHYPPSHLSLEAGND
RSLHOALEKARAGLGNDPARSLIKWPHFYGISDMSFLGAAASGSOVVSNTPLSLVD
RPSENALRPTVNLGPWGREFHQFERVHVPYAFVLPVLVSEIARTFLGDDRRHD"
complement(4701..6353)
/genes="SMB20698"
complement(4701..6353)
/genes="SMB20698"
/function="CELL PROCESSES; Transport of small molecules"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative transport protein, similar to E. coli
multidrug resistance protein EmrB"
/protein_id="CAC49805.1"
/db_xref="GI:15141293"
/db_xref="GOA:Q92TU0"
/db_xref="SPTREMBL:Q92TU0"
/translacion="MTIATPSTRDISMAGLLVAGIVLATEIASTVLSLGRGDI
IGDYATPDEFAMLDVGYTFLKIGFMAASWMSRDPNVLVGSTLAMGACGTAAM
TARLDLVALRMTIQSGSTLVGGQAIIFTFPHSRQPLQAFAMGSAVAPATAP
ALCQLLDQSMTWIFSIIPLAAGAALLADGVAQRAORFPDFWIGFSLIGSAL
FCFTYVFSQSRWDWEEPIRLWLAVIGTATLLAFLGQOVAKGQGLPDTLPFETEDF
CFATYVSFVAGALFCSAFLIPSPAVSLAFTPDAGOLLPSGALFICALIARFLM
QLRNPFPVAFPGILNINWAMMWSGTSSEGADGMKAILLRNGLGLFLSLTLL
AFSRNSRNLASGILNFTGRQJGLIGVSLTLEHNVSQALILLRNGLGLFLSLTLL
ADLT'TAALLTAKGMDAAASGLLGRAVAGOSTVIAFDNFAFNLGLLFLVTAAP
VLVGIKIGLARYAKARAEGRAGAAKPSGHHTLPKRVPAS"
complement(6350..7447)

```



```

CDS
/gene="SMB20699"
complement(6350..7447)
/gene="SMB20699"
/function="CELL PROCESSES: Transport of large molecules;
protein, peptide secretion"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="putative protein secretion protein, HlyD family,
similar to E. coli EmrX"

Query Match 87.0%; Score 17.4; DB 1; Length 189333;
Best Local Similarity 94.7%; Pred. NO. 1.1e-03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCCGCCAGCCCCACA 19
Db 80571 CGAGGCCGCCAGCCCCGCA 80589

RESULT 23
AF274573/c
LOCUS AF274573 189436 bp DNA linear PRI 01-JAN-2003
DEFINITION Homo sapiens chromosome X clone RP11-316K19, RP11-316M19, RP4-687A5
map q28, complete sequence.
ACCESSION AF274573
VERSION AF274573.4 GI:27452915
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189436)
AUTHORS Galgoczy,P., Schilhabel,M., Rosenthal,A. and Platzer,M.
TITLE Chromosome X genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189436)
AUTHORS Platzer,M.
TITLE Direct Submission
SUBMITTER Direct Submission
SUBMITTED 02-JUN-2000 Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
3 (bases 1 to 189436)
AUTHORS Platzer,M.
TITLE Direct Submission
SUBMITTED 22-JUN-2002 Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
4 (bases 1 to 189436)
AUTHORS Platzer,M.
TITLE Direct Submission
SUBMITTED 01-JAN-2003 Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
On Jan 1, 2003 this sequence version replaced gi:21539160.
----- Genome Center
Center: Institute of Molecular Biotechnology
Center code: IMB
Web site: http://genome.imb-jena.de/
Contact: gscj-submit@genome.imb-jena.de
----- Project Information
Center project name: x131+
Center clone name: RP11-316K19, RP4-687A5
----- Summary Statistics
Sequencing vector: M13, pUC18, 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 183203 bases at least Q40
Consensus quality: 189383 bases at least Q30
Consensus quality: 189436 bases at least Q20
Quality coverage: 18.73x
-----
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate

```

```

chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest.
-----
Neighboring sequence information:
This entry is overlapped by QC-11G7, LLOXNC01-P0142,
ICRFXC104-M0525, LLOXNC01-C1233, QC-3C1, LLOXNC01-B1439 and covers
RP11-316K19, LLOXNC01-D0132, RP4-687A5, LLOXNC01-223D9 entirely.
RP11-316M19 is identical to RP11-316K19.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----

```

```

FEATURES
             Location/Qualifiers
     source   1..189436
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /chromosome="X"
              /map="q28"
              /clone="RP11-316K19, RP11-316M19, RP4-687A5"
              1..2112
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /clone="QC-11G"
              /clone_lib="Q12-derived Xq27.3-Xqter cosmid library"
              /notes="overlapping clone"
     source   1..30055
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /clone="LLOXNC01-P014"
              /clone_lib="Lawrence Livermore human X cosmid library"
              /notes="overlapping clone"
     source   1..185184
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /clone="RP11-316K19"
              /clone_lib="RPCI human BAC 11"
              13446..39763
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /clone="LLOXNC01-D0132"
              /clone_lib="Lawrence Livermore human X cosmid library"
              86923..189436
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /clone="RP4-687A5"
              /clone_lib="RPCI human PAC 4"
              136581..174193
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /clone="LLOXNC01-223D9"
              /clone_lib="Lawrence Livermore human X cosmid library"
              163186..189436
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /clone="ICRFXC104-M052"

```

```

/clone_lib="RCRF human X specific cosmid library no.104"
/note="overlapping clone"
165363. .189436
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="LLOXNC01-C123"
/note="overlapping clone"
/clone_lib="Lawrence Livermore human X cosmid library"
174464. .189436
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Qc-3C"
/note="Qc-3C"
/clone_lib="Q12-derived Xq27.3-Xqter cosmid library"
177285. .189436
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="LLOXNC01-B143"
/note="overlapping clone"
674
/note="C substituted in clone: Qc-11G7"
/replace="T"
1677
/note="C substituted in clone: LLOXNC01-PO142"
/replace="T"
2522
/note="A substituted in clone: LLOXNC01-PO142"
/replace="G"
4257
/note="C substituted in clone: LLOXNC01-PO142"
/replace="T"
5340. .5341
/note="deleted in clone: LLOXNC01-PO142"
/replace="G"
5825
/note="G substituted in clone: LLOXNC01-PO142"
/replace="C"
6442
/note="deleted in clone: RP11-316K19"
/replace="A"
7042
/note="C substituted in clone: LLOXNC01-PO142"
/replace="T"
8639
/note="G substituted in clone: LLOXNC01-PO142"
/replace="A"
9064
/note="deleted in clone: RP11-316K19"
/replace="A"
9546. .9549
/note="deleted in clone: RP11-316K19"
/replace="A"
9595. .9598
/note="deleted in clone: RP11-316K19"
/replace="A"
10095. .10098
/note="deleted in clone: RP11-316K19"
/replace="A"
11036
/note="T substituted in clone: LLOXNC01-PO142"
/replace="C"
12039
/note="C substituted in clone: LLOXNC01-PO142"
/replace="A"
12684
/note="C substituted in clone: LLOXNC01-PO142"
/replace="G"
13364
/note="C substituted in clone: LLOXNC01-PO142"

```

```

/replace="T"
14099. .14115
/note="deleted in clone: RP11-316K19"
/replace="A"
15637
variation
variation
Query Match 87.0%; Score 17.4; DB 9; Length 189436;
Best Local Similarity 94.7%; Pred. No. 1.le+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CGAGGCCGCCAGCCGCCACA 19
|||||
DB 180044 CGAGGTGCCAGCCGCCACA 180026
|||||
RESULT 24
AC100872/c
LOCUS AC100872 195563 bp DNA linear PRI 01-MAR-2002
DEFINITION Homo sapiens chromosome 8, clone RP11-973F15, complete sequence.
ACCESSION AC100872.2 GI:19033627
VERSION AC100872.2
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 195563)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,B.
TITLE Homo sapiens chromosome 8, clone RP11-973F15
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 195563)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,B., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepeil,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamarez,R., Landers,T., Lechoczy,J., Levine,R., Liu,G.,
MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norb,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollata,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thumann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 195563)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,B., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepeil,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamarez,R.,
Landers,T., Lechoczy,J., Levine,R., Liu,G., MacLean,C.,
MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollata,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,

```



```

Best Local Similarity 94.7%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCAGCCCAACAA 20
Db 63244 GAGGCGCACAGCCCAACAA 63226

RESULT 25
AC124859/c
LOCUS AC124859
DEFINITION Homo sapiens chromosome 5 clone RP11-876M24, WORKING DRAFT
SEQUENCE, 8 unordered pieces.
ACCESSION AC124859
VERSION AC124859.1 GI:21450430
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 197000)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 197000)
DOE Joint Genome Institute.
Direct Submission
Submitted (18-JUN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1631200
Center clone name: RPCI-11_876M24

Summary Statistics
Consensus quality: 186113 bases at least Q40
Consensus quality: 191620 bases at least Q30
Consensus quality: 193891 bases at least Q20
Estimated insert size: 219000; agarose-ep estimation
Estimated insert size: 7.49 in Q20 bases; sum-of-contigs estimation
Quality coverage: 8.36 in Q20 bases; sum-of-contigs estimation
Quality coverage: 8.36 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2446: contig of 2446 bp in length
2447 2546: gap of unknown length
2547 7929: contig of 5383 bp in length
7930 7930: gap of unknown length
8030 21479: contig of 13450 bp in length
21480 21579: gap of unknown length
21580 34877: contig of 13298 bp in length
34878 34977: gap of unknown length
34978 59976: contig of 24999 bp in length
59977 60076: gap of unknown length
60077 91893: contig of 31817 bp in length
91894 91993: gap of unknown length
91994 137246: contig of 45252 bp in length
137246 137346: gap of unknown length
137346 197000: contig of 59655 bp in length.
Location/Qualifiers
1. .197000
/organism="Homo sapiens"
/mol_type="genomic DNA"

```

```

/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-876M24"
/clone_lib="RPCI human BAC library 11"

ORIGIN

Query Match 87.0%; Score 17.4; DB 2; Length 197000;
Best Local Similarity 94.7%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCAGCCCAACAA 20
Db 48167 GAGGCGCCAGCCCAACAA 48149

RESULT 26
AC023224/c
LOCUS AC023224
DEFINITION Homo sapiens chromosome 17 clone RP11-420J13 map 17, WORKING DRAFT
SEQUENCE, 21 unordered pieces.
ACCESSION AC023224
VERSION AC023224.3 GI:7145009
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 20437)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 17, clone RP11-420J13
Unpublished
2 (bases 1 to 20437)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Becker, R., Bieda, F.,
Boguski, M., Brown, A., Brown, A., Burkett, S., Castle, A.,
Choe, Y., Collinge, M., Collins, S., Collins, S., Collymore, A., Cooke, P.,
DeArnell, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
Ferrel, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,
Lander, E., Leach, J., Levine, R., Liu, C., Liu, G., Locke, K.,
MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPherson, R., Meidrid, J., Meneses, L., Morrow, J., Naylor, J.,
Norman, C., O'Connor, T., O'Donnell, P., Oliver, T., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced gi:7139828.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Smit, A.P.A. & Green, P. (1996-1997)

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L6679
Center clone name: 420_J13
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 189335 bases at least Q40
Consensus quality: 195441 bases at least Q30
Consensus quality: 196187 bases at least Q20
Insert size: 205000; agarose-ep
Insert size: 200437; sum-of-contigs

```

Quality coverage: 4.7 in Q20 bases; agarose-fp  
 Quality coverage: 4.9 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 21 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1
* 1811 1910: contig of 1810 bp in length
* 1910: gap of 100 bp
* 1911 5018: contig of 3108 bp in length
* 5018: gap of 100 bp
* 5019 5118: gap of 100 bp
* 5118: gap of 100 bp
* 5119 7959: contig of 2841 bp in length
* 7959: gap of 100 bp
* 7960 8059: gap of 100 bp
* 8059: gap of 100 bp
* 8060 11595: contig of 3536 bp in length
* 11595: gap of 100 bp
* 11596 11795: gap of 100 bp
* 11795: gap of 100 bp
* 11796 14271: contig of 2476 bp in length
* 14271: gap of 100 bp
* 14272 22386: contig of 8015 bp in length
* 22386: gap of 100 bp
* 22387 22486: gap of 100 bp
* 22486: gap of 100 bp
* 22487 29746: contig of 7260 bp in length
* 29746: gap of 100 bp
* 29747 29846: gap of 100 bp
* 29846: gap of 100 bp
* 29847 37497: contig of 7651 bp in length
* 37497: gap of 100 bp
* 37498 37597: gap of 100 bp
* 37597: gap of 100 bp
* 37598 43589: contig of 5992 bp in length
* 43589: gap of 100 bp
* 43590 43689: gap of 100 bp
* 43689: gap of 100 bp
* 43690 50270: contig of 6581 bp in length
* 50270: gap of 100 bp
* 50271 50370: gap of 100 bp
* 50370: gap of 100 bp
* 50371 57887: contig of 7517 bp in length
* 57887: gap of 100 bp
* 57888 57987: gap of 100 bp
* 57987: gap of 100 bp
* 57988 65412: contig of 7425 bp in length
* 65412: gap of 100 bp
* 65413 75598: contig of 10186 bp in length
* 75598: gap of 100 bp
* 75599 75798: gap of 100 bp
* 75798: gap of 100 bp
* 75799 85662: contig of 9864 bp in length
* 85662: gap of 100 bp
* 85663 85762: gap of 100 bp
* 85762: gap of 100 bp
* 85763 97087: contig of 11325 bp in length
* 97087: gap of 100 bp
* 97088 97187: gap of 100 bp
* 97187: gap of 100 bp
* 97188 113019: contig of 15832 bp in length
* 113019: gap of 100 bp
* 113020 130955: contig of 17836 bp in length
* 130955: gap of 100 bp
* 130956 131055: gap of 100 bp
* 131055: gap of 100 bp
* 131056 148692: contig of 17637 bp in length
* 148692: gap of 100 bp
* 148693 148792: gap of 100 bp
* 148792: gap of 100 bp
* 148793 165067: contig of 16275 bp in length
* 165067: gap of 100 bp
* 165068 165167: gap of 100 bp
* 165167: gap of 100 bp
* 165168 185335: contig of 20168 bp in length
* 185335: gap of 100 bp
* 185336 185435: gap of 100 bp
* 185435: gap of 100 bp
* 185436 202437: contig of 17002 bp in length.
  
```

## FEATURES

## source

```

1..202437
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-420J13"
/clone_lib="RPC1-11 Human Male BAC"
1..1810
/note="assembly_fragment"
1911..5018
/note="assembly_fragment"
5119..7959
/note="assembly_fragment"
8060..11695
/note="assembly_fragment"
11796..14271
/note="assembly_fragment"
14372..22386
/note="assembly_fragment"
  
```

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

```

misc_feature 22487..23746
/note="assembly_fragment"
misc_feature 29847..37497
/note="assembly_fragment"
misc_feature 37598..43589
/note="assembly_fragment"
clone_end:17
vector_side:left"
43690..50270
/note="assembly_fragment"
50371..57887
/note="assembly_fragment"
57988..65412
/note="assembly_fragment"
65513..75698
/note="assembly_fragment"
75799..85662
/note="assembly_fragment"
85763..97087
/note="assembly_fragment"
97188..113019
/note="assembly_fragment"
113120..130955
/note="assembly_fragment"
131056..148692
/note="assembly_fragment"
148793..165067
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
165168..185335
/note="assembly_fragment"
185436..202437
/note="assembly_fragment"
  
```

## ORIGIN

Query Match 87.0%; Score 17.4; DB 2; Length 202437;  
 Best Local Similarity 94.7%; Pred. No. 1.1e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCGAGCCCCACA 19

Db 171824 CGAGGTGCCGAGCCCCACA 171806

## RESULT 27

RME603643 RME603643 260050 bp DNA linear BCT 05-JUL-2002  
 LOCUS Rhizobium meliloti (Sinorhizobium meliloti) 1021 complete pSYMB;  
 DEFINITION segment 2/6.

ACCESSION AL603643 AL591985

VERSION AL603643.1 GI:15140164

## KEYWORDS

SOURCE Sinorhizobium meliloti (Rhizobium meliloti)

## ORGANISM

Sinorhizobium meliloti  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.

## REFERENCE

## AUTHORS

1  
 Finan, T.M., Weidner, S., Wong, K., Buhrmester, J., Chain, P.,  
 Vorholter, F.J., Hernandez-Lucas, I., Becker, A., Cowie, A., Gouzy, J.,  
 Golding, B. and Puhler, A.  
 The complete sequence of the 1,683-kb pSYMB megaplasmid from the  
 N2-fixing endosymbiont Sinorhizobium meliloti

Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9889-9894 (2001)

## JOURNAL

## MEDLINE

## PUBMED

## REMARK

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REMARK

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

COMMENT	Submitted on behalf of Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany and Department of Biology, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4K1 Canada mailto:Stefan.Weidner@Genetik.Uni-Bielefeld.DE PEXO, pSymB.
FEATURES	Location/Qualifiers 1..260050 /organism="Sinorhizobium meliloti" /mol_type="genomic DNA" /strain="1021" /db_xref="taxon:382" /plasmid="pSymB" complement(120..281) /gene="SMB20302" complement(120..281) /genes="SMB20302" /function="MISCELLANEOUS; Hypothetical/Global homology" /note="Product confidence : putative Gene name confidence : hypothetical" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="CONSERVED HYPOTHETICAL PROTEIN" /protein_id="CAC48692.1" /db_xref="GI:15140165" /db_xref="GOA:Q92WN7" /db_xref="SPTREMBL:Q92WN7" /translation="MSVLISILITFLVILVLYLVNRLPLDGRTRKQIVQVIVIIIGIL SLLRYLAVP" 478..705 /gene="SMB20303" 478..705 /genes="SMB20303" /function="MISCELLANEOUS; Unknown" /note="Product confidence : hypothetical Gene name confidence : hypothetical" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="HYPOTHETICAL PROTEIN" /protein_id="CAC48693.1" /db_xref="GI:15140166" /db_xref="GOA:Q92WN6" /db_xref="SPTREMBL:Q92WN6" /translation="MSVRFSDLAIDDDIILMLRRVHVYACSSRELEPESEHGAELGKM LPLHYNQVRSLEALMQVLTAPEAPPEAG" complement(746..1804) /note="this element seems to be partial or inactive" /insertion_seq="ISRM2011-2/ISRM11 OR SMB21616" complement(762..1367) /gene="TRM2011-2C OR SMB20304" complement(762..1367) /gene="TRM2011-2C OR SMB20304" /function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related functions" /note="Product confidence : probable Gene name confidence : putative" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="probable ISRM2011-2 transposase protein, C-terminal portion" /protein_id="CAC48694.1" /db_xref="GI:15140167" /db_xref="GOA:Q92527" /db_xref="SPTREMBL:Q92527" /translation="KKTVALSERPPDVARHARWLKHCPIIDPARLVFIDETWTKTN MAPLRGWARGRLVGYAPFGWNTMTFVALRADRVSAPIIDGPIINGERFRLYVQQ VLVPELKAQDIIVLNDIGSHKGCEINAAIRKAGARLFLPKYISPLNPIELKPAKIKH WLREQAQRSDAIHDLRHILOAVTPQECAYFKEAGYERA" 1002..1373 /gene="TRM2011-2 OR SMB20306"
repeat_region	
gene	
CDS	
1002..1373 /gene="TRM2011-2 OR SMB20306" /function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related functions" /note="Product confidence : putative Gene name confidence : putative" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="putative TRM2011-2 protein" /protein_id="CAC48695.1" /db_xref="GI:15140168" /db_xref="GOA:Q52888" /db_xref="SPTREMBL:Q52888" /translation="MGAEIVQDDVDVAGFFERHQNLIDIAEAFADRAIEDKGSADAV GPECGDGHGVPMAGGVGAHQSFARERPAPQRHVLRPLGLIDENAGGINSGLVLPQ SCPVAGDVGLPLPFAFDQRLFL" complement(1301..1708) /gene="TRM2011-2N OR SMB20305" complement(1301..1708) /gene="TRM2011-2N OR SMB20305" /function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related functions" /note="Product confidence : probable Gene name confidence : putative" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="probable ISRM2011-2 transposase protein, N-terminal portion" /protein_id="CAC48696.1" /db_xref="GI:15140169" /db_xref="GOA:Q926A8" /db_xref="SPTREMBL:Q926A8" /translation="MARPSNDLRERVVDVAVTGESLSCRAAAKRGIGISTALDWVRR FRETSAAPQMGHGRKLSGPHRAWLLCRCRDFTLHGLVAELSERGLKVDYRAV WTFVHEGLSYKKRWSPPANGSGPTSPATGHDG" complement(1801..2448) /gene="SMB20307" complement(1801..2448) /gene="SMB20307" /function="MISCELLANEOUS; Hypothetical/Partial homology" /note="Product confidence : putative Gene name confidence : hypothetical" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="putative dihydroxyacetone kinase protein" /protein_id="CAC48697.1" /db_xref="GI:15140170" /db_xref="GOA:Q92WN5" /db_xref="SPTREMBL:Q92WN5" /translation="MMSESAVRRLIEVFEAISAEIARDHLSLSDGAIGDADHGITM AIGFSAVTGLARTDAGSLGVTAAASAFNLVAGSTGPLYATAPRAAQALEGR DTLIDACSLIIQAVASGIGERGKGQSDKTMDLWLPADAAQAQAAAKSTAAFWA DVTAAENGASATRSNVATKGRARVGRSLGHMDPFAASAVLIRAMAQTILREG" complement(2527..4614) /gene="SMB20312" complement(2527..4614) /gene="SMB20312" /function="MISCELLANEOUS; Not classified regulator" /note="Product confidence : putative Gene name confidence : hypothetical" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="putative transcriptional regulator protein" /protein_id="CAC48698.1" /db_xref="GI:15140171" /db_xref="GOA:Q92WN4" /db_xref="SPTREMBL:Q92WN4" /translation="MTIKTSPARKGAANSRGTIDASDAIPLRYGDDPYVWACHLYYED GMTQGEIAEAMGVSRATVNSYLADAREKGI VNI SI EPARLASLTI AQELKRHFGLTDC	
gene	

source	1. .14033 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="cosmid 358b7" 1833. .4356 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="cosmid 316g12" 39209. .78044 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="cosmid 399e4" 75764. .110196 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="cosmid 312e8" 95097. .126950 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="cosmid 438f12" 110851. .157562 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="cosmid 390e6" 142264. .185653 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="cosmid 305c8" 184733. .220367 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="cosmid 385e7" 187817. .233498 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="cosmid 380f5" 232370. .276380 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="cosmid 313f4" 248926. .279011 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="cosmid 425c2" 275910. .279011 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="cosmid 315g5" <1. .1932 /note="H335887" complement(1294. .1450) /notes="MIR; RepeatMasker predicted 23/10/2000" rpt_family="SINE/MIR" 1833. .39308 /note="H3316G12" 2453. .2501 /notes="MER5A; RepeatMasker predicted 23/10/2000" rpt_family="DNA/MER1_type" 2517. .2817 repeat_region
--------	---

```
misc_feature
repeat_region
repeat_region
```

```

/note="AluJo; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
2819..2926
/note="MER5A; RepeatMasker predicted
23/10/2000"
/rpt_family="DNA/MER1_type"
2839..4637
/note="CpG island; BAIAP3"
/evidence=not_experimental
3444..3502
/note="GC rich; RepeatMasker predicted
23/10/2000"
/rpt_family="Low_complexity"
3559..19342
/gene="BAIAP3"
/note="synonym: BAP3"
4319..4339
/note="GC rich; RepeatMasker predicted
23/10/2000"
/rpt_family="Low_complexity"
join(4565..4659,8443..8583,8805..8892,9067..9147,
9399..9506,10825..10884,11032..11098,11197..11392,
11725..11769,11889..11966,12084..12186,12430..12637,
12622..12717,12816..12919,13296..13416,13925..14049,
14130..14220,14372..14428,14504..14601,14701..14752,
14872..15002,15152..15275,15650..15793,15889..15981,
16073..16264,16345..16413,16508..16587,16750..16858,
17201..17272,17448..17528,17604..17724,17797..17920,
17999..18204,18282..19342)
/gene="BAP3"
/note="Similar to Unigene cluster Hs.101516"
join(4565..4659,8443..8583,8805..8892,9067..9147,
9399..9506,10825..10884,11032..11098,11197..11392,
11725..11769,11889..11966,12084..12186,12430..12637,
12622..12717,12816..12919,13296..13416,13925..14049,
14130..14220,14372..14428,14504..14601,14701..14752,
14872..15002,15152..15275,15650..15793,15889..15981,
16073..16264,16345..16413,16508..16587,16750..16858,
17201..17272,17448..17528,17604..17724,17797..17920,
17999..18204,18282..18385)
/gene="BAP3"
/codon_start=1
/product="BAI associated protein 3"
/protein_id="AAK6125.1"
/db_xref="GI:14336746"
/db_xref="MIM:604009"
/db_xref="SPTREMBL:O94812"
/translation="MRRGAAPAFPGDHLGTAIGFAGAIMRSRSPAMSTLIDKS
SVLRQVQCPSEFRSTDPQASADPOEATGAWKPGDGVFFAHMLMKKGEGQ
GLPCLVPLRSGSPAPPEVDPSLGLALAPEVEMLYEALYTVLYAGTMGDVD
DEALLSTLQVFGTSEETHEALERVKAKPYALKVSVWRAKNLLAKDPGFSPP
YCMGLIPASDAPTEPRAQKRGFRGKGGPLPAKCIQVTEVKSSTLNPVKRH
FLFTIEDYTDHLMDHDDSVLSVEACRKNELVGLKMGIVKQIVKRSANST
AGPTEDHDDFLGCLNTPVPVAVGVDFWFLKLEPRSSRSGELSTPAATILCLHQAQNSL
TAMSGRSGFLSHLLLSHLLRHEAGAEENPSWRSGLSTPAATILCLHQAQNSL
PQLAVLHWQVSSRHOCTDYSYLLGLLEDQMOAHEEAPSLPQOEESLADSLAF
SEFGLQLRLQRDFYPATNSTAVHRIELLLKCLQLFQPSFEICPESELNADIAA
ALKKNRWYDRILNDSFPQPGQRLPGLVLADAVDDLQRCYSVYASLHSLIN
VDVFTLTPQLERLIVAEAWLVEELSPMTLVEASGLFELTLADQLRFDWSIFGR
DSRLALAGHAPLPAVLKQVQLVQVLRQAKRLQAVMDLPEVDASSHSSAAFA
GLCLSHIQELVRLWLPDPAQAGLQGLQGVCEATLFYTELLRKVDTPQGAAGA
VSEALCVNLNVELVRKAAQALGLAWPEGATGPEGLPRLPCLSCQALDLDQREA
HTVTAHLTKMVQDIRKVOHLSIPDSINDEAVFLMLDKELALLNVLKVNKL
SRVLEALLQQLKIQVALLNSPDSIDFYSRFTLEALVSPFAEQGLPLBSLRD
GSYKRLKELRHKCSTRECTEQYLDKLKORTLEQNRFLSVCHYEAREFLAVE
VLHADDLPLDANGSLDFVLELGGPFLPFLVRSQRTQVKTIRTLHPVDFYFYSV
AEACRRCAACVLTVMDDHJSTNDFAEAGALGGVGTAVRQVGGGAGARQGPVTL
LCRPQAQVRSALRLEGTSTSEAKGFVKKLEKCEKMEADP"
5744..5876
/note="MIR; RepeatMasker predicted
23/10/2000"

```



/organism="Sinorhizobium meliloti"
/mol\_type="genomic DNA"
/strain="1021"
/db\_xref="taxon:382"
70..813
/gene="gst11 OR SMC00916"
70..813
/gene="gst11 OR SMC00916"
/EC\_number="2.5.1.18"
/function="miscellaneous"
/notes="Product confidence : putative
Gene name confidence : putative
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon\_start=1
/evidence=not experimental
/transl\_table=11
/product="PUTATIVE GLUTATHIONE S-TRANSFERASE PROTEIN"
/protein\_id="CAC45361.1"
/db\_xref="GI:15073720"
/db\_xref="GOA:Q92RR5"
/db\_xref="SPTREMBL:Q92RR5"
/translation="MADLSRPITARWPAKNPDIITQLYSLPTNGVKVSI ALEELGLA
YEAHRSPTNEQKSPFELSINPNRIRIPALDPNGPDGKPIGLFSGAILLYLAETG
KLUPDAARRYETLCWVGMQGIQGFHFFFAADKVNPNPYVQRYDEAKR
LUGVLEAREGKWLWGGNTPTSPTRIPGKRGSSWVAFRCSTIRASRSWTGWI
AAWHLRFRWRSSRSRSRSLPAGD"
588..779
/gene="SMC00915"
588..779
/gene="SMC00915"
/function="miscellaneous; hypothetical/partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon\_start=1
/evidence=not experimental
/transl\_table=11
/product="HYPOTHETICAL PROTEIN"
/protein\_id="CAC45362.1"
/db\_xref="GI:15073721"
/db\_xref="SPTREMBL:Q92KL5"
/translation="WADGGEYTIADIATYPWEGARKFYGAEVLYDKSPFNWDVWV
DRGLRPAARQKGMEIPRKE"
867..2084
/gene="SMC00914"
867..2084
/gene="SMC00914"
/function="small molecule metabolism"
/notes="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon\_start=1
/evidence=not experimental
/transl\_table=11
/product="PUTATIVE OXIDOREDUCTASE PROTEIN"
/protein\_id="CAC45363.1"
/db\_xref="GI:15073722"
/db\_xref="GOA:Q92RR4"
/db\_xref="SPTREMBL:Q92RR4"
/translation="MSGLVVGQQAFAALVAKLRALKDNRPIITVAASLPYQRP
PLSKYLRENTLRLTRPAWAEHIDIRLSTTVTRDRLAKQVALSDGSMLTTE
TLATATGTPRLPAAGDLAGVVFVDFDADRLAEQPGSRVIRVGGYIGLEA
AAVATSGLEADRIQIRVASAATSIVREIRHSGVIDRERTGLHRLIGNG
RVTAELSDGVPVIVIGVIGADALAEAGIETANGIVVDHSRGTSDPTIVAM
GDCAVLPMWGMRIRESVONAVDOAEAAVLAGGTDPYDPKPMFSDQVDVKLIQIAG
FLGLHDETFLVQGORQGSVSVYFRQQLIAVDINDAKAYVTGCKLLEAGATPDRLL
LADQVDLKALIV"

repeat\_region
complement(2114..2572)
/notes="Sm-2 OR SMC04876
REPEAT SM-2
Predicted by Homology"
/evidence=not experimental
complement(2585..4222)
/gene="groEL1 OR SMC00913"
complement(2585..4222)
/gene="groEL1 OR SMC00913"
/function="cell processes; chaperoning"
/notes="Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon\_start=1
/evidence=not experimental
/transl\_table=11
/product="60 KD CHAPERONIN A PROTEIN"
/protein\_id="CAC45364.1"
/db\_xref="GI:15073723"
/db\_xref="GOA:P35469"
/db\_xref="SWISS-PROT:P35469"
/translation="MAKEVKFGRSAREKMLRGVDILADAVKVTLPKGRNVVIDKSP
GAPRITKDGVSVAKEIELEDKFENMGQWVREASKNTNDIAGGTTATVLAQIVRE
GAKAVAGMFMOLKRGIDJAVAEVVDLLAKKAKINTSDEVAQVGTISANGKQIGL
DIAMQKNGNEGVITVEAKTAELEVEGQFDRGYLSPTVTPNPERMVADLEDA
FIILHEKLSNLOAMLPLVAVVQTKPLIIAEDVEGEALTLVWNLKGGKIIAAV
KAPFGRRKAMLEDAITLGTGTIVIGEDLGIKLESVTLMLGRKAVSIGKENTIVD
GAGQSDIERGVAIKAOIETTSDYDREKLOERLAKLAGSVAIVRGGATEVEVKEK
KDRIDDAINATRAVQEGIVPGGVALLESVKITVKGENDDQDAGVIVRRALQSPA
RQIVENAGDEASIVVCKLEKNTDDRGYNATGEYGDMTIANGIIDPVKVVRTALQDRA
SVASLLITTEAMIAELPKDAPANPGMGMGGMGMDMM"
complement(4298..4594)
/gene="groES1 OR SMC00912"
complement(4298..4594)
/gene="groES1 OR SMC00912"
/function="cell processes; chaperoning"
/notes="Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon\_start=1
/evidence=not experimental
/transl\_table=11
/product="10 KD CHAPERONIN A PROTEIN"
/protein\_id="CAC45365.1"
/db\_xref="GI:15073724"
/db\_xref="GOA:P35473"
/db\_xref="SWISS-PROT:P35473"
/translation="MASTNFRPLHDRVVRVVESEKTKGIIIPDTAKEKPOEGEIV
AVGSGARDSGSKVPLDVKAGDRILFGKSGTEVKINGEDLLIMKEADIMGVIG"
4863..5690
/gene="SMC00911"
4863..5690
/gene="SMC00911"
/function="miscellaneous; hypothetical/global homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Framed"
/codon\_start=1
/evidence=not experimental
/transl\_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein\_id="CAC45366.1"
/db\_xref="GI:15073725"
/db\_xref="SPTREMBL:Q92RR3"
/translation="MAFNSISSEKLARLVGTFKAPVIVDVRDDDDFAADHHLVPGSF
HRGHVHPQWAEYRGQVTVLCKHGKLGSEAAWLRHAGAAAILDGGIEMWRAAG
LPIVPAFNPSPDAGSGVVMVTRSRPKIDRIACPLIRFVDFPRARFLVTPAEVAV
GERFGATPDIDEGVFWSHRGECTPTMTYKESGLSPALEHVARIVRGADTDRRELEP
EAAGLLAVSLGLSRMYADLELEAGMTIYDALYWARDATQEKHDWVSHGKGRKP"
6004..6852
/gene="SMC00910"
6004..6852

```

/gene="SMC00910"
/function="miscellaneous; hypothetical/global homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE GLUTATHIONE S-TRANSFERASE PROTEIN"
/protein_id="CAC45361.1"
/db_xref="GI:15073720"
/db_xref="GOA:Q92RR2"

Query Match      87.0%; Score 17.4; DB 1; Length 286550;
Best Local Similarity 94.7%; Pred. No. 1.le+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCAGCCGCCACA 19
Db 25466 CGAGGCGCCAGCCGCCCA 25484

RESULT 30
SME591785/c
LOCUS SME591785 286550 bp DNA linear BCT 05-JUL-2002
DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 4/12.
ACCESSION AL591785 AL591688
VERSION AL591785.1 GI:15073719
KEYWORDS Sinorhizobium meliloti (Rhizobium meliloti)
SOURCE Sinorhizobium meliloti
ORGANISM Sinorhizobium meliloti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
Boisard, P., Becker, A., Boutry, M., Cadieu, E., Dreano, S., Gloux, S.,
Godrie, T., Goffeau, A., Kahn, D., Kiss, E., Lelaure, V., Masuy, D.,
Pohl, T., Portetelle, D., Puehler, A., Purnelle, B., Ramsperger, U.,
Renard, C., Thebault, P., Vandenbol, M., Weidner, S. and Galibert, F.
Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
MEDLINE 21396507
PUBMED 11481430
REFERENCE 2 (bases 1 to 286550)
AUTHORS Gouzy, J.
TITLE Direct Submission
SUBMITTED (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
EU Consortium
MELILO EU Consortium:
Laboratoire de Biologie Molculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr
http://sequence.toulouse.inra.fr/meliloti.html.
Location/Qualifiers
1. 286550
/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
70. .813
/gene="gst11 OR SMC00916"
70. .813

FEATURES
source
1. 286550
/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
70. .813
/gene="gst11 OR SMC00916"
70. .813

/gene="gst11 OR SMC00916"
/EC number="2.5.1.18"
/function="miscellaneous"
/notes="Product confidence : putative
Gene name confidence : putative
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE GLUTATHIONE S-TRANSFERASE PROTEIN"
/protein_id="CAC45361.1"
/db_xref="GI:15073720"
/db_xref="GOA:Q92RR5"
/db_xref="SPTREMBL:Q92RR5"
/translation="MADLASPPITASPAKPNPDIOLYSLPTFNGVKVSIABEELGLA
YEHRISPDFTNEOKSPFELSLAPNGIPALIDPNGEDGKIGLFSGAILLYLBKGTG
KLIPADARRYETLCWFMFGMGIGNFGQGHFFKFAADKANNPPYQVRYDEAKR
LLGVLEARLEGQWLMGGRNTPSPTSRPIPGMKGSGSWAPFCSTIRASRMSWTGWI
AAWHALRPRAWRSRARSRSRSLPAD"
588. .779
/gene="SMC00915"
588. .779
/gene="SMC00915"
/function="miscellaneous; hypothetical/partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEITICAL PROTEIN"
/protein_id="CAC45362.1"
/db_xref="GI:15073721"
/db_xref="SPTREMBL:Q92KL5"
/translation="MADGGEYTIATYTPWVGARKFYGGAEVLDYKSFNVMWV
DRGLARPAQKQMEIPRKE"
867. .2084
/gene="SMC00914"
867. .2084
/gene="SMC00914"
/function="small molecule metabolism"
/notes="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE OXIDOREDUCTASE PROTEIN"
/protein_id="CAC45363.1"
/db_xref="GI:15073722"
/db_xref="GOA:Q92RR4"
/db_xref="SPTREMBL:Q92RR4"
/translation="MSCRLVVGQGAAPALVAKLRALKDMRPITVVAEASLPYORP
PLSKYLLREMTLDLLVREAVVAEHEIDRLSTVTRVDRLAKOVALSDGSLTYE
TLAPATPRLPAAVGGDLGAVVVRDRDRLAEEMQGRRLVVGSGYIGLEA
AAVATSGLEVTEIEMADRIQORVASAATSAI VREIHRSGVDIRERTGHLRLGDN
RVTAAELSDGSVIPDIVIGVANDALAHAGIETANGI VVDSHGRVSDFTIVAM
GICHLVPMWGNRIRESQVNAQAEAAVLAGTDPDPKPFWSDDQYDKLVNAG
FGCHFDTELVROGQGSVSVVPRQKLIADINDAKAYVTGKKLLEAGATPDRLL
LADQVLDKALLV"
complement (2114. .2572)
/notes="Sm-2 OR SMC04876
REPEAT SM-2
Predicted by Homology"
/evidence=not_experimental
complement (2585. .4222)
/gene="groEL1 OR SMC00913"

```

CDS

```
complement(2585..4222)
/genes="groEL1 OR SMC00913"
/function="cell processes; chaperoning"
/note="predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="60 KD CHAPERONIN A PROTEIN"
/protein_id="CAC45364.1"
/db_xref="GI:15073723"
/db_xref="GOA:P35469"
/db_xref="SWISS-PROT:P35469"
/translation="MAAEVVFGRSAREKMLRGVLDLADAVKVTILGPKRNVWIDKSP
GAPRITKQCVSAKEIELEDPENNMAQWREVAASKTNDIAGDQTTTATVLAQAIIVRE
GAKAVAGMNPDLKRGIDLAVAEVVDKLLAKKINTSDEVAOVGTTISANGKQIGL
DIAEAMQVKGEGVITVEAKTAETELVWGMQPDRCGLSPFYFTNPEKMWALIEDA
FILLHEKLSNLQMLPVLVAEVVQIGKELLIAEDVEGALATLVNKLKGLKIAAV
KAPGGRKRAVLEDAIILTGTVISEDLGKLESVTLDMLGRAKKVSIITKENTIVD
GAGQSDTEGRAVQIAQITBETTSIDYREKLQERLAKLAGGAVIRVGGATEVEVKEX
KDIRIDALNATRAAQEGTIPGGVALLRSVKITVKGENDDQDAGVNVIRALQSPA
ROIIVENAGDEASIVVGTLEKNTDGFYNATGEGYDGMIANGLIIDPVKVVETALQDAA
SVASLLITTEAMIAELPKKADAPMGGMGGMDDM"
complement(4298..4594)
/genes="groES1 OR SMC00912"
/complement(4298..4594)
/genes="groES1 OR SMC00912"
/function="cell processes; chaperoning"
/note="predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="10 KD CHAPERONIN A PROTEIN"
/protein_id="CAC45365.1"
/db_xref="GI:15073724"
/db_xref="GOA:P35473"
/db_xref="SWISS-PROT:P35473"
/translation="WASNFRLPHDRVVVRVESEKTKGGIIPDTAKKPPQGEIV
AVGSGARDESGKVPFLVDYKAGDRILFGKMSTEVKINGEDLLIMKEADIMGVIG"
4863..5690
/genes="SMC00911"
4863..5690
/genes="SMC00911"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC45366.1"
/db_xref="GI:15073725"
/db_xref="SPTREMBL:Q92RR3"
/translation="WASFNSISSEKLARLVCTPKAPVIVDVRDDDDFAADHHLVPGSF
IRGHVHVQWAEETRGVGVVVICVHGKLGSEGAARLHAGAAEILDDGLEAWRAAG
LPLVVASLNSDPAGSVVWVTRSPKIDRIACPLWIRFVDPFARFLYVTPAEVEAV
GERGATFDIEDGVFWSHRGECTFDTNWKEFGLSFPALHVARIVRGADTDRELEP
EAGALLVASLGLSRMYADDLQLEAGMTIYDALYRWARDAQEKXHDWVSHGKGKRP"
6004..6852
/genes="SMC00910"
6004..6852
/genes="SMC00910"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
```

gene

CDS

```
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC45367.1"
/db_xref="GI:15073726"
/db_xref="GOA:Q92RR2"

Query Match      87.0%   Score 17.4; DB 1; Length 286550;
Best Local Similarity 94.7%   Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 CGAGGGCCCCCAGCCCCACA 19
Db      284901 CGAGGGCCCCCAGCCCCGCA 284883
|||||
|||||

Search completed: June 20, 2004, 11:42:05
Job time : 569.069 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:46:26 ; Search time 114.921 Seconds  
(without alignments)  
739.327 Million cell updates/sec

Title: US-10-624-714-11

Perfect score: 20  
Sequence: 1 cgaggccgcccccacaa 20

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

N Geneseq\_29Jan04:\*

1: geneseqn1980s:\*

2: geneseqn1990s:\*

3: geneseqn2000s:\*

4: geneseqn2001as:\*

5: geneseqn2001bs:\*

6: geneseqn2002as:\*

7: geneseqn2003as:\*

8: geneseqn2003bs:\*

9: geneseqn2003cs:\*

10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	95.0	110000	4	AAI99682_25
2	19	95.0	110000	4	AAI99682_26
3	19	95.0	110000	4	AAI99683_25
4	16.8	84.0	410	2	AXX20696
5	16.8	84.0	993	6	ABQ75352
6	16.8	84.0	1002	6	AAQD39099
7	16.8	84.0	1830	7	ACA52540
8	16.8	84.0	9851	9	ADC86884
9	16.8	84.0	19186	2	AXX20620
10	16.8	84.0	21501	9	ADC35539
11	16.8	84.0	28000	9	ADD00956
12	16.8	84.0	37113	6	ABQ75121
13	16.4	82.0	354	3	AAAC10189
14	16.4	82.0	3145	3	AAAC59241
15	16.4	82.0	3287	6	AAAS99902
16	16.4	82.0	3345	6	ABN59766
17	16.4	82.0	19951	7	ABZ24596
18	16	80.0	3417	6	ABV78124
19	16	80.0	3417	6	ABZ35700
20	16	80.0	3417	6	ABX09943
21	16	80.0	3417	6	ABL91665
22	16	80.0	3845	2	AAQ47064
23	16	80.0	3845	6	ABZ34892

6	ABV94119	Breast ca
7	ABX08795	Angiogene
8	ABX12538	CDNA enco
9	ABX63148	Human CDN
10	ACH03957	Human MET
11	AAC90316	XI16619 CD
12	ABL61947	Colon ade
13	ABL68365	Kidney ca
14	ABL61948	Colon ade
15	ABL61948	Colon ade
16	ABL61948	Colon ade
17	ABL61948	Colon ade
18	ABL61948	Colon ade
19	ABL61948	Colon ade
20	ABL61948	Colon ade
21	ABL61948	Colon ade
22	ABL61948	Colon ade
23	ABL61948	Colon ade
24	ABL61948	Colon ade
25	ABL61948	Colon ade
26	ABL61948	Colon ade
27	ABL61948	Colon ade
28	ABL61948	Colon ade
29	ABL61948	Colon ade
30	ABL61948	Colon ade
31	ABL61948	Colon ade
32	ABL61948	Colon ade
33	ABL61948	Colon ade
34	ABL61948	Colon ade
35	ABL61948	Colon ade
36	ABL61948	Colon ade
37	ABL61948	Colon ade
38	ABL61948	Colon ade
39	ABL61948	Colon ade
40	ABL61948	Colon ade
41	ABL61948	Colon ade
42	ABL61948	Colon ade
43	ABL61948	Colon ade
44	ABL61948	Colon ade
45	ABL61948	Colon ade
46	ABL61948	Colon ade
47	ABL61948	Colon ade
48	ABL61948	Colon ade
49	ABL61948	Colon ade
50	ABL61948	Colon ade
51	ABL61948	Colon ade
52	ABL61948	Colon ade
53	ABL61948	Colon ade
54	ABL61948	Colon ade
55	ABL61948	Colon ade
56	ABL61948	Colon ade
57	ABL61948	Colon ade
58	ABL61948	Colon ade
59	ABL61948	Colon ade
60	ABL61948	Colon ade
61	ABL61948	Colon ade
62	ABL61948	Colon ade
63	ABL61948	Colon ade
64	ABL61948	Colon ade
65	ABL61948	Colon ade
66	ABL61948	Colon ade
67	ABL61948	Colon ade
68	ABL61948	Colon ade
69	ABL61948	Colon ade
70	ABL61948	Colon ade
71	ABL61948	Colon ade
72	ABL61948	Colon ade
73	ABL61948	Colon ade
74	ABL61948	Colon ade
75	ABL61948	Colon ade
76	ABL61948	Colon ade
77	ABL61948	Colon ade
78	ABL61948	Colon ade
79	ABL61948	Colon ade
80	ABL61948	Colon ade
81	ABL61948	Colon ade
82	ABL61948	Colon ade
83	ABL61948	Colon ade
84	ABL61948	Colon ade
85	ABL61948	Colon ade
86	ABL61948	Colon ade
87	ABL61948	Colon ade
88	ABL61948	Colon ade
89	ABL61948	Colon ade
90	ABL61948	Colon ade
91	ABL61948	Colon ade
92	ABL61948	Colon ade
93	ABL61948	Colon ade
94	ABL61948	Colon ade
95	ABL61948	Colon ade
96	ABL61948	Colon ade

97	15.4	77.0	858	6	ABK99585	Abk99585 Rat proba	c 170	15.2	76.0	448	2	AAQ59153	AaQ59153 Human bra
98	15.4	77.0	858	7	ACD07310	AcD07310 Human muc	c 171	15.2	76.0	475	7	ACC44118	Acc44118 Herpes si
99	15.4	77.0	882	2	AX57407	Ax57407 Rat U3 ge	c 172	15.2	76.0	523	4	AAH21061	AaH21061 Parallel
100	15.4	77.0	984	6	AX57499	Ax57499 Human G-p	c 173	15.2	76.0	665	7	ABT40532	AbT40532 Toxicity
101	15.4	77.0	984	6	AB242916	Ab242916 Human GFC	c 174	15.2	76.0	903	7	ACF39423	Acf39423 Mycobacte
102	15.4	77.0	984	6	AAI99826	AaI99826 Human GFC	c 175	15.2	76.0	906	7	ACA40706	AcA40706 Prokaryot
103	15.4	77.0	990	6	AAI99826	AaI99826 Human G p	c 176	15.2	76.0	963	7	ACA39806	AcA39806 Prokaryot
104	15.4	77.0	990	6	ABV72908	AbV72908 Human G p	c 177	15.2	76.0	1126	6	AB564603	Ab564603 Human cDN
105	15.4	77.0	990	6	ABV72908	AbV72908 Human G p	c 178	15.2	76.0	1146	7	ACC78413	Acc78413 E. desert
106	15.4	77.0	993	6	AB154463	Ab154463 Human GFC	c 179	15.2	76.0	1449	9	ADG64479	AdG64479 Human inh
107	15.4	77.0	993	6	ABK991022	AbK991022 Human G-p	c 180	15.2	76.0	1796	9	ADG64479	AdG64479 Human inh
108	15.4	77.0	993	6	ABK991022	AbK991022 Human GFC	c 181	15.2	76.0	1806	6	ABK77885	AbK77885 Nucleotid
109	15.4	77.0	993	6	ABK991022	AbK991022 Human GFC	c 182	15.2	76.0	1806	6	ABK77885	AbK77885 Nucleotid
110	15.4	77.0	993	6	ABK991022	AbK991022 Human GFC	c 183	15.2	76.0	1806	6	ABK77885	AbK77885 Nucleotid
111	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 184	15.2	76.0	2426	1	AAH50490	AaH50490 Sequence
112	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 185	15.2	76.0	2426	1	AAH50490	AaH50490 Sequence
113	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 186	15.2	76.0	2784	9	ACA37694	AcA37694 Prokaryot
114	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 187	15.2	76.0	2839	9	ACA37694	AcA37694 Prokaryot
115	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 188	15.2	76.0	2839	9	ACA37694	AcA37694 Prokaryot
116	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 189	15.2	76.0	2839	9	ACA37694	AcA37694 Prokaryot
117	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 190	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
118	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 191	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
119	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 192	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
120	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 193	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
121	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 194	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
122	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 195	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
123	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 196	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
124	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 197	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
125	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 198	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
126	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 199	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
127	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 200	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
128	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 201	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
129	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 202	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
130	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 203	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
131	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 204	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
132	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 205	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
133	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 206	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
134	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 207	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
135	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 208	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
136	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 209	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
137	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 210	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
138	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 211	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
139	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 212	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
140	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 213	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
141	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 214	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
142	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 215	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
143	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 216	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
144	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 217	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
145	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 218	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
146	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 219	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
147	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 220	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
148	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 221	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
149	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 222	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
150	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 223	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
151	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 224	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
152	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 225	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
153	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 226	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
154	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 227	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
155	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 228	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
156	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 229	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
157	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 230	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
158	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 231	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
159	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 232	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
160	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 233	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
161	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 234	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
162	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 235	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
163	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 236	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
164	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 237	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
165	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 238	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
166	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 239	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
167	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 240	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
168	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 241	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN
169	15.4	77.0	1038	7	AB242594	Ab242594 Human GFC	c 242	15.2	76.0	3016	4	AAH16088	AaH16088 Human cDN

243	14.8	74.0	238	3	AAC28264	Human sec
C 244	14.8	74.0	240	4	AS36090	Human car
C 245	14.8	74.0	240	9	AdE46784	Human car
C 246	14.8	74.0	247	7	AdA48755	Human acy
C 247	14.8	74.0	253	4	AS36088	Human car
C 248	14.8	74.0	253	4	AS36087	Human car
C 249	14.8	74.0	253	9	AdE46782	Human car
C 250	14.8	74.0	253	9	AdE46781	Human car
C 251	14.8	74.0	316	2	AdT22911	Human gen
C 252	14.8	74.0	332	2	AdQ60504	Human bra
C 253	14.8	74.0	332	6	AdL85098	Human ova
C 254	14.8	74.0	353	6	AdL84737	Human ova
C 255	14.8	74.0	389	4	AdL23824	Human bre
C 256	14.8	74.0	391	6	AdL87076	Human ova
C 257	14.8	74.0	422	4	AdI92169	Human pol
C 258	14.8	74.0	426	4	AdI23034	Colon tum
C 259	14.8	74.0	426	7	AdB233220	Human col
C 260	14.8	74.0	438	4	AdI14987	Probe #49
C 261	14.8	74.0	438	4	AdA56720	Human foe
C 262	14.8	74.0	438	4	AdI36333	Probe #50
C 263	14.8	74.0	438	4	AdA46177	Human bre
C 264	14.8	74.0	438	4	AdA26339	Probe #48
C 265	14.8	74.0	438	4	AdA30378	Human bon
C 266	14.8	74.0	438	4	AdA04850	Human bra
C 267	14.8	74.0	438	4	AdS30017	Human liv
C 268	14.8	74.0	438	5	AdI04752	Probe #47
C 269	14.8	74.0	438	6	AdS04976	Human gen
C 270	14.8	74.0	474	2	AdV86659	EST clone
C 271	14.8	74.0	500	4	AdI14971	Human bre
C 272	14.8	74.0	501	8	AdC31903	Human end
C 273	14.8	74.0	502	4	AdI98936	Human exc
C 274	14.8	74.0	502	5	AdI63286	Human kid
C 275	14.8	74.0	519	4	AdH13057	Human cdn
C 276	14.8	74.0	520	6	AdN84933	Human glu
C 277	14.8	74.0	532	7	AdB236649	Human gen
C 278	14.8	74.0	532	3	AdZ50398	Human end
C 279	14.8	74.0	565	4	AdI17173	Probe #71
C 280	14.8	74.0	565	4	AdA61700	Human foe
C 281	14.8	74.0	565	4	AdI41612	Probe #10
C 282	14.8	74.0	565	4	AdA29343	Probe #78
C 283	14.8	74.0	565	4	AdK35896	Human bon
C 284	14.8	74.0	565	4	AdK10002	Human bra
C 285	14.8	74.0	565	4	AdS35605	Human liv
C 286	14.8	74.0	565	6	AdS10119	Human gen
C 287	14.8	74.0	591	6	AdH10551	Human cdn
C 288	14.8	74.0	591	6	AdR84931	Human glu
C 289	14.8	74.0	596	4	AdH69170	Human cer
C 290	14.8	74.0	598	6	AdK16211	Human lun
C 291	14.8	74.0	598	9	AdB95474	Human lun
C 292	14.8	74.0	610	3	AdH31084	Human col
C 293	14.8	74.0	616	6	AdL93019	Rat metas
C 294	14.8	74.0	620	3	AdA16138	Human col
C 295	14.8	74.0	620	3	AdC03957	Human sec
C 296	14.8	74.0	624	6	AdK33069	DNA encod
C 297	14.8	74.0	638	4	AdA24470	Human ova
C 298	14.8	74.0	638	5	AdH83074	Human ova
C 299	14.8	74.0	640	2	AdX97692	Extended
C 300	14.8	74.0	648	7	AdB36503	Human GEN

ALIGNMENTS

RESULT 1  
AAI99682\_25  
Continuation (26 of 45) of AAI99682 from base 2500001 (Mycobacterium tuberculosis strain  
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682  
WP Fragment Name Begin End  
WP AAI99682\_00 1 110000  
WP AAI99682\_01 100001 210000  
WP AAI99682\_02 200001 310000  
WP AAI99682\_03 300001 410000  
WP AAI99682\_04 400001 510000  
WP AAI99682\_05 500001 610000  
WP AAI99682\_06 600001 710000  
WP AAI99682\_07 700001 810000  
WP AAI99682\_08 800001 910000  
WP AAI99682\_09 900001 1010000  
WP AAI99682\_10 1000001 1110000  
WP AAI99682\_11 1100001 1210000  
WP AAI99682\_12 1200001 1310000  
WP AAI99682\_13 1300001 1410000  
WP AAI99682\_14 1400001 1510000  
WP AAI99682\_15 1500001 1610000  
WP AAI99682\_16 1600001 1710000  
WP AAI99682\_17 1700001 1810000

WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

Query Match 95.0%; Score 19; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 GAGGCGCCGAGCCCAAA 20  
Db 104135 GAGGCGCCGAGCCCAAA 104153  
|||||

RESULT 2  
AAI99682\_26  
Continuation (27 of 45) of AAI99682 from base 2600001 (Mycobacterium tuberculosis strain  
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682  
WP Fragment Name Begin End  
WP AAI99682\_00 1 110000  
WP AAI99682\_01 100001 210000  
WP AAI99682\_02 200001 310000  
WP AAI99682\_03 300001 410000  
WP AAI99682\_04 400001 510000  
WP AAI99682\_05 500001 610000  
WP AAI99682\_06 600001 710000  
WP AAI99682\_07 700001 810000  
WP AAI99682\_08 800001 910000  
WP AAI99682\_09 900001 1010000  
WP AAI99682\_10 1000001 1110000  
WP AAI99682\_11 1100001 1210000  
WP AAI99682\_12 1200001 1310000  
WP AAI99682\_13 1300001 1410000  
WP AAI99682\_14 1400001 1510000  
WP AAI99682\_15 1500001 1610000  
WP AAI99682\_16 1600001 1710000  
WP AAI99682\_17 1700001 1810000





```
Db      131 CGAGCGCCCAATCCCAAA 112
RESULT 5
ABQ75352/c
ID   ABQ75352 standard; cDNA; 993 BP.
XX
AC   ABQ75352;
XX
DT      05-NOV-2002 (first entry)
XX
DE      Human lung specific nucleic acid sequence SEQ ID NO:91.
XX
KW      Human; lung; lung specific nucleic acid; LSNA; lung specific protein; LSP;
KW      LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;
KW      squamous cell carcinoma; gene; ss.
XX
OS      Homo sapiens.
XX
PN      WO200264788-A2.
XX
PD      22-AUG-2002.
XX
PF      20-NOV-2001; 2001WO-US045080.
XX
PR      20-NOV-2000; 2000US-0252054P.
XX
PA      (DIAD-) DIADEXUS INC.
XX
PI      Macina RA, Recipon H, Chen S, Sun Y, Liu C;
XX
WI      WPI; 2002-657601/70.
XX
New lung specific nucleic acid useful in gene therapy or as vaccines for
PT      treating lung cancer (e.g. squamous cell carcinoma) or non-cancerous lung
PT      diseases, as well as for diagnosing, monitoring or staging these
PT      diseases.
XX
PS      Claim 1; Page 206-207; 282pp; English.
XX
CC      The present invention describes an isolated lung specific nucleic acid
CC      (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid
CC      sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965;
CC      (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp);
CC      given in ABQ75262 to ABQ75376; (c) selectively hybridises to (a) or (b);
CC      or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific
CC      protein (LSP) sequences have cytostatic activity and can be used in gene
CC      therapy and vaccines. LSNA and LSPs are useful for diagnosing and
CC      monitoring the presence and metastases of lung cancer in a patient. An
CC      antibody that specifically binds to an LSP can be used for determining
CC      the presence of an LSP in a sample, as well as for treating a patient
CC      with lung cancer, particularly by inducing an immune response against the
CC      lung cancer cell expressing the LSNA or LSPs. In particular, these LSNA
CC      and LSPs are useful for identifying, diagnosing, monitoring, staging,
CC      imaging and treating lung cancer (e.g. squamous cell carcinoma) and non-
CC      cancerous disease states in lung
XX
SQ      Sequence 993 BP; 210 A; 225 C; 312 G; 246 T; 0 U; 0 Other;
Query Match      84.0%; Score 16.8; DB 6; Length 993;
Best Local Similarity 90.0%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy      1 CGAGCGCCCAATCCCAAA 20
Db      485 CGCGGCCCCAGCCCAAA 466
RESULT 6
AAD39099/c
ID   AAD39099 standard; cDNA; 1002 BP.
XX
AC   AAD39099;
XX
DT      19-JUN-2003 (first entry)
XX
DE      Human lung specific nucleic acid (LSNA) #6.
XX
KW      Human; lung-specific nucleic acid; LSNA; lung-specific protein; LSP;
KW      lung cancer; non-cancerous disease; gene therapy; transgenic animal;
KW      transgenic; engineered lung tissue; cytostatic; gene; ss.
XX
OS      Homo sapiens.
XX
PN      WO200240673-A2.
XX
PD      23-MAY-2002.
XX
PF      20-NOV-2001; 2001WO-US045180.
XX
PR      20-NOV-2000; 2000US-0252055P.
XX
PR      22-NOV-2000; 2000US-0252496P.
XX
PA      (DIAD-) DIADEXUS INC.
XX
PI      Macina RA, Recipon H, Chen S, Sun Y, Liu C, Turner LR;
XX
WI      WPI; 2002-471623/50.
XX
P-PSDB; AAE24326.
XX
New lung-specific nucleic acids, useful for identifying, diagnosing,
PT      monitoring, staging, imaging, and treating lung cancer and non-cancerous
PT      disease states in lung tissues.
XX
PS      Claim 1; Page 150-151; 187pp; English.
XX
CC      The invention relates to new lung-specific nucleic acids (LSNA) and lung-
CC      specific proteins (LSP). The lung-specific nucleic acids, polypeptides
CC      and compositions comprising them are useful: for identifying, diagnosing,
CC      monitoring, staging, imaging, and treating lung cancer and non-cancerous
CC      disease states in lung tissue; for identifying lung tissue; for
CC      monitoring, identifying and/or designing agonists and antagonists of the
CC      polypeptides; in gene therapy; in producing transgenic animals and cells;
CC      for producing engineered lung tissue for treatment and research; and as
CC      elements in an array or computer program for pattern recognition of lung
CC      disorders. The nucleic acids may be used as hybridisation probes to
CC      detect, characterise and quantify hybridising nucleic acids in and
CC      isolate hybridising nucleic acids from, both genomic and transcript-
CC      derived nucleic acid samples. The present sequence is human lung-specific
CC      cDNA
XX
SQ      Sequence 1002 BP; 214 A; 241 C; 334 G; 213 T; 0 U; 0 Other;
Query Match      84.0%; Score 16.8; DB 6; Length 1002;
Best Local Similarity 90.0%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy      1 CGAGCGCCCAATCCCAAA 20
Db      559 CGCGGCCCCAGCCCAAA 540
RESULT 7
ACA52540/c
ID   ACA52540 standard; DNA; 1830 BP.
XX
AC   ACA52540;
XX
DT      19-JUN-2003 (first entry)
```



XX Treponema pallidum.  
XX OS WO9859034-A2.  
XX PN 30-DEC-1998.  
XX PD 23-JUN-1998; 98WO-US013041.  
XX PF 24-JUN-1997; 97US-0050667P.  
XX PR (HUMA-) HUMAN GENOME SCI INC.  
XX PA Fraser CM;  
XX PI WPI; 1999-081273/07.  
XX DR  
XX PT New isolated Treponema pallidum nucleic acids - used to develop products  
XX PT for the detection, diagnosis, characterisation, prevention and therapy of  
XX PT T. pallidum infections, particularly syphilis.  
XX PS Claim 1; Page 736-747; 1150pp; English.  
XX CC AA220500-21243 represent polynucleotide sequences from the genome of  
XX CC Treponema pallidum. The sequences can be used for detection, diagnosis,  
XX CC characterisation, prevention and therapy for T. pallidum infections,  
XX CC particularly syphilis. They can also be used for detecting diseases  
XX CC related to Borrelia infections in animals, and for the production of  
XX CC biosynthetic products such as enzymes  
XX CC  
XX SQ Sequence 19186 BP; 4705 A; 5450 C; 4881 G; 4131 T; 0 U; 19 Other;  
Query Match 84.0%; Score 16.8; DB 2; Length 19196;  
Best Local Similarity 90.0%; Pred. No. 5.3e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CGAGGCGCCGAGCCCAAA 20  
DB 9003 CGAGGCGCCGAGCCCAAA 9022  
RESULT 10  
ADC35539/C  
ID ADC35539 standard; DNA; 21501 BP.  
XX AC ADC35539;  
XX DT 18-DEC-2003 (first entry)  
XX DE Human CD81/TAPA-1 partial genomic sequence.  
XX KW Antisense; ds; human; CD81; TAPA-1; tetraspanin; viral infection;  
XX KW cocaine addiction; autoimmune disorder; antiinflammatory; antibacterial;  
XX KW virucide; antiparasitic; inflammatory disorder; parasitic infection;  
XX KW bacterial infection.  
XX OS Homo sapiens.  
XX PN US2003113914-A1.  
XX PD 19-JUN-2003.  
XX PF 10-DEC-2001; 2001US-00006430.  
XX PR 10-DEC-2001; 2001US-00006430.  
XX PA (ISIS-) ISIS PHARM INC.  
XX PI Graham MJ, Dobie K;  
XX DR WPI; 2003-810907/76.  
XX PT Novel compound hybridizing with nucleic acid molecule encoding CD81 and

PT inhibiting the expression of CD81, useful for treating infections and  
PT disease associated with expression of CD81 such as inflammation disorder.  
XX Example 15; SEQ ID NO 11; 55pp; English.  
XX The invention relates to a compound (antisense oligonucleotide)  
CC hybridising with the eighth nucleobase portion of an active site on a  
CC nucleic acid molecule encoding CD81 (also known as TAPA-1, a tetraspanin)  
CC and inhibiting the expression of CD81. Also included is a composition  
CC comprising the antisense oligonucleotide and a carrier or a diluent. The  
CC antisense oligonucleotide is useful for inhibiting the expression of CD81  
CC in cells or tissues. The antisense oligonucleotide is also useful for  
CC treating infections preferably viral, bacterial and parasitic and  
CC diseases such as inflammatory disorders and autoimmune disorders. The  
CC disease or condition is characterised by chemical dependency (e.g.  
CC cocaine addiction). The present sequence is a Human CD81/TAPA-1 partial  
CC genomic sequence.  
XX SQ Sequence 21501 BP; 3604 A; 6408 C; 7119 G; 4370 T; 0 U; 0 Other;  
Query Match 84.0%; Score 16.8; DB 9; Length 21501;  
Best Local Similarity 90.0%; Pred. No. 5.3e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CGAGGCGCCGAGCCCAAA 20  
DB 7509 CGAGGCGCCGAGCCCAAA 7490  
RESULT 11  
ADD00956  
ID ADD00956 standard; DNA; 28000 BP.  
XX AC ADD00956;  
XX DT 01-JAN-2004 (first entry)  
XX DE Human Jagged 2 encoding DNA SEQ ID NO:11.  
XX KW apoptosis; Jagged 2 inhibitor; cytostatic; hyperproliferative disorder;  
XX KW human; ds; gene; chromosome 14.  
XX OS Homo sapiens.  
XX PN WO2003077848-A2.  
XX PD 25-SEP-2003.  
XX PF 10-MAR-2003; 2003WO-US007340.  
XX PR 12-MAR-2002; 2002US-00096399.  
XX PA (ISIS-) ISIS PHARM INC.  
XX PI Koller E, Shapard PJ;  
XX DR WPI; 2003-756943/71.  
XX DR P-PSDB; ADD01037.  
XX PT Inducing apoptosis in a cell or animal for treating a subject having a  
XX PT condition associated with insufficient apoptosis by administering to a  
XX PT cell or animal a Jagged 2 inhibitor to reduce Jagged 2 levels or  
XX PT activity.  
XX PS Example 15; SEQ ID NO 11; 148pp; English.  
XX The present invention describes a method for inducing apoptosis in a cell  
CC or animal comprising administering to a cell or animal a Jagged 2  
CC inhibitor to reduce Jagged 2 levels or activity. Also described: (1)  
CC treating a subject having a disease or condition associated with  
CC insufficient apoptosis by administration of a Jagged 2 inhibitor; (2) a  
CC pharmaceutical composition comprising a Jagged 2 inhibitor and another  
CC active ingredient for inducing apoptosis; and (3) a kit comprising a

CC Jagged 2 inhibitor and instructions for using the Jagged 2 inhibitor in  
 CC the induction of apoptosis. The Jagged 2 inhibitor has cytostatic  
 CC activity. The method can be used for inducing apoptosis in a cell or  
 CC animal for treating a subject having a disease or condition associated  
 CC with insufficient apoptosis, e.g., hyperproliferative disorder. The  
 CC present sequence encodes the human Jagged 2 protein, which is used in an  
 CC example from the present invention. Human Jagged 2 is located on  
 CC chromosome 14.  
 XX  
 SQ Sequence 28000 BP; 4121 A; 8288 C; 10292 G; 5299 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 9; Length 28000;  
 Best Local Similarity 90.0%; Pred. No. 5.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGAGGGCGCCAGCCCAACAA 20  
 Db 26057 CCAGCGCCGAGCCCAACAA 26076

RESULT 12  
 ABQ75121/c  
 ID ABQ75121 standard; DNA; 37113 BP.

XX AC ABQ75121;

XX DT 04-NOV-2002 (first entry)

XX DE Human CD81 gene SEQ ID NO:11.

XX KW Human; p15 region; chromosome 11; tumour growth; infertility; cytostatic;  
 XX hepatitis C virus infection; antiinfertility; gene therapy; CD 81;  
 XX cluster of differentiation antigen 81; gene; ds.

XX OS Homo sapiens.

XX FH Location/Qualifiers

FT CDS 10471..29787  
 FT /tag= a  
 FT /product= "cluster of differentiation antigen 81 (CD 81)  
 FT protein"

FT /note= "contains introns"

FT exon 10471..10536

FT /tag= b

FT /number= 1

FT intron 10537..23332

FT /tag= c

FT /number= 1

FT exon 23333..23446

FT /tag= d

FT /number= 2

FT intron 23447..27014

FT /tag= e

FT /number= 2

FT exon 27015..27113

FT /tag= f

FT /number= 3

FT intron 27114..27892

FT /tag= e

FT /number= 3

FT exon 27893..27964

FT /tag= f

FT /number= 4

FT intron 27963..28333

FT /tag= e

FT /number= 4

FT exon 28334..28441

FT /tag= f

FT /number= 5

FT intron 28442..28789

FT /tag= e

FT /number= 5

FT exon 28790..28891

FT /tag= f  
 FT /number= 6  
 FT intron 28892..29548  
 FT /tag= e  
 FT /number= 6  
 FT exon 29549..29635  
 FT /tag= f  
 FT /number= 7  
 FT intron 29636..29724  
 FT /tag= e  
 FT /number= 7  
 FT exon 29725..29784  
 FT /tag= f  
 FT /number= 8

XX WO200261085-A2.

XX PD 08-AUG-2002.

XX 31-OCT-2001; 2001WO-US045381.

XX 31-OCT-2000; 2000US-0244705P.

XX (RYAN/) RYAN J W.

XX PI Ryan JW;

XX DR WPI; 2002-619251/66.

XX P-PSDB; ABP52845.

XX New genes obtainable from the p15 region of human chromosome 11 (e.g.  
 human achaete-scute homolog 2), useful in gene therapy, particularly for  
 preventing or treating tumor growth, infertility or hepatitis C virus  
 infection.

XX Claim 1; Fig 2; 94pp; English.

XX The present invention describes an isolated genomic polynucleotide (I),  
 CC which is obtainable from the p15 region of human chromosome 11.  
 CC Specifically described are the human achaete-scute homologue 2 (HASH2),  
 CC SMS3, tumour suppressing sub-transferable candidate 6 (TSSC6), ribosomal  
 CC protein L26 (RIBO26), cluster of differentiation antigen 81 (CD 81) and  
 CC tumour suppressing sub-transferable candidate 4 (TSSC4) genes as given in  
 CC ABQ75117 to ABQ75122, encoding the proteins given in ABP52841 to  
 CC ABP52846. Also described is an isolated polynucleotide obtainable from  
 CC the p15 region of chromosome 11 having the sequence, which comprises any  
 CC of two nucleotide sequences given in ABQ75123 and ABQ75124. (I) has  
 CC cytostatic and antiinfertility activities and can be used in gene  
 CC therapy. The HASH2, SMS3, TSSC6, RIBO26, CD 81 or TSSC4 polynucleotide  
 CC sequences can be used for preventing, treating or ameliorating a medical  
 CC condition, or for manufacturing a gene therapy for the prevention,  
 CC treatment or amelioration of such medical conditions. An antisense  
 CC oligonucleotide or mimetic of (I) can also be useful for manufacturing a  
 CC medicament for the prevention, treatment or amelioration of these medical  
 CC conditions. These conditions include tumour growth, infertility, or  
 CC hepatitis C virus infection

SQ Sequence 37113 BP; 6829 A; 10985 C; 11249 G; 8050 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 6; Length 37113;

Best Local Similarity 90.0%; Pred. No. 5.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGAGGGCGCCAGCCCAACAA 20

Db 17233 CAAGGGCGCCGAGCCCAACAA 17214

RESULT 13

AAC10189

ID AAC10189 standard; cDNA; 354 BP.

XX AC AAC10189;

```
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 14264.
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EP1033401-A2.
XX 06-SEP-2000.
XX 21-FEB-2000; 2000EP-00200610.
XX 26-FEB-1999; 99US-0122487P.
XX (GEST ) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX Claim 1; SEQ ID NO 14264; 71bp + Sequence Listing; English.
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
XX ESTs are derived from mRNAs with intact 5' ends and can therefore be used
XX to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors
XX
XX Sequence 354 BP; 58 A; 155 C; 85 G; 53 T; 0 U; 3 Other;
XX
Query Match 82.0%; Score 16.4; DB 3; Length 354;
Best Local Similarity 94.4%; Pred. No. 8.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CGAGGCGCCCGAGCCCCAC 18
DB 306 CGAGGCGCCCGAGCCCCAC 323
RESULT 14
AAC59241
ID AAC59241 standard; cDNA; 3145 BP.
XX AAC59241;
XX 29-JAN-2001 (first entry)
XX Human secreted protein cDNA sequence #27.
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein; ss.
XX Homo sapiens.
OS
```

```
XX WO200055199-A1.
XX 21-SEP-2000.
XX 09-MAR-2000; 2000WO-US006014.
XX 12-MAR-1999; 99US-0124095P.
XX 11-JUN-1999; 99US-0138598P.
XX 03-DEC-1999; 99US-0168665P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Komatsoulis G;
XX WPI; 2000-572359/53.
XX P-PSDB; AAB27820.
XX Isolated nucleic acid molecule encoding a human secreted protein is used
XX in preventing, treating or ameliorating a medical condition.
XX Claim 1; Page 355-356; 433pp; English.
XX The invention relate to the isolation of genes AAC59215-C59261 encoding
XX 47 human secreted proteins AAB27794-B27840. The genes can be used to
XX generate fusion proteins by linking to the gene for the human
XX immunoglobulin G Fc portion for increasing the stability of the fusion
XX protein as compared to the human protein only. The genes and proteins are
XX useful for preventing, ameliorating or treating medical conditions, e.g.
XX by protein or gene therapy. The genes are isolated from a range of human
XX tissues disclosed in the specification. The nucleic acids, proteins,
XX antibodies and (ant)agonists are useful in the diagnosis, treatment and
XX prevention of: (a) cancer, e.g. breast and ovarian cancer, and other
XX cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal
XX tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's
XX disease, allergies, autoimmune haemolytic anaemia, autoimmune
XX thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
XX rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders
XX such as myocardial ischaemias; (d) wound healing; (e) neurological
XX diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases
XX such as viral, bacterial, fungal and parasitic infections
XX
XX Sequence 3145 BP; 806 A; 1006 C; 745 G; 581 T; 0 U; 7 Other;
XX
Query Match 82.0%; Score 16.4; DB 3; Length 3145;
Best Local Similarity 94.4%; Pred. No. 8e-02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CGAGGCGCCCGAGCCCCAC 18
DB 37 CGAGGCGCCCGAGCCCCAC 54
RESULT 15
AAS99902
ID AAS99902 standard; cDNA; 3287 BP.
XX AAS99902;
XX 12-MAR-2002 (first entry)
XX Polynucleotide encoding human cytoskeleton-associated protein #13.
XX Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder; ss;
XX cell proliferative disorder; inflammatory disorder; prion disease;
XX vesicle trafficking disorder; gastrointestinal disorder; muscle disorder;
XX neurological disorder; cell motility disorder; reproductive disorder;
XX spinal cord disease; central nervous system disorder; mental disorder;
XX gene therapy; cancer.
XX Homo sapiens.
XX WO200185942-A2.
XX
```

XX 15-NOV-2001.  
 XX 03-MAY-2001; 2001WO-US014355.  
 XX 05-MAY-2000; 2000US-0201960P.  
 XX 08-MAY-2000; 2000US-0202729P.  
 XX 05-JUN-2000; 2000US-0209705P.  
 XX 07-JUN-2000; 2000US-0210149P.  
 XX 21-JUN-2000; 2000US-0213215P.  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hillman JL;  
 PI Azimzai Y, Lal P, Yao MG, Bandman O, Burford N, Batra S, Kearney L;  
 PI Policky JL;  
 XX WPI; 2002-062248/08.  
 DR P-PSDB; AAU74342.  
 XX New cytoskeleton-associated proteins and polynucleotides, useful for  
 PT diagnosing, preventing and treating cell proliferative, autoimmune,  
 PT inflammatory, neurological, cell motility, reproductive and muscle  
 PT disorders.  
 XX Claim 5; Page 178; 194pp; English.  
 XX The invention relates to human cytoskeleton-associated polypeptides  
 CC (CYSPK) and their associated polynucleotide sequences. The sequences are  
 CC useful in the treatment of disorders associated with overexpression or  
 CC underexpression of CYSPK in a patient. The disorders include cell  
 CC proliferative disorders (such as cancer, actinic keratosis,  
 CC arteriosclerosis, cirrhosis, hepatitis and psoriasis),  
 CC autoimmune/inflammatory disorders (such as, asthma, atherosclerosis,  
 CC osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus  
 CC and anaemia), vesicle trafficking disorders (such as  
 CC hypercholesterolaemia, diabetes insipidus, grave's disease and goitre),  
 CC gastrointestinal disorders, prion diseases, neurological disorders (such  
 CC as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,  
 CC Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis  
 CC and other motor neuron disorders), cell motility disorders, reproductive  
 CC disorders (such as endometriosis and polycystic ovary syndrome), muscle  
 CC disorders (such as myocarditis, migraine, hypertension, hypoglycaemia,  
 CC myocardial infarction, epilepsy and muscular dystrophy), spinal cord  
 CC diseases, central nervous system disorders (such as Down syndrome and  
 CC cerebral palsy) and mental disorders (such as anxiety and schizophrenia).  
 CC Sequences AAS93890-AAS9923 represent cDNA molecules encoding human CYSPK  
 CC of the invention  
 XX SQ Sequence 3287 BP; 795 A; 1095 C; 788 G; 609 T; 0 U; 0 Other;  
 Query Match 82.0%; Score 16.4; DB 6; Length 3287;  
 Best Local Similarity 94.4%; Pred. No. 8e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CGAGGCCGCCAGCCCCAC 18  
 DB 220 CGAGGCCGCCAGCCCCAC 237  
 RESULT 16  
 ID ABN59766 standard; cDNA; 3345 BP.  
 XX AC ABN59766;  
 XX 28-JUN-2002 (first entry)  
 XX Novel human coding sequence SEQ ID NO: 177.  
 DE Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;  
 XX Human; antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
 KW neuroprotective; antiparkinsonian; protein therapy; EST;  
 KW

XX expressed sequence tag; gene; ss.  
 OS Homo sapiens.  
 XX WO200222660-A2.  
 XX 21-MAR-2002.  
 XX 10-SEP-2001; 2001WO-US026015.  
 XX 11-SEP-2000; 2000US-00659671.  
 XX (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Dmanac RT;  
 XX WPI; 2002-292408/33.  
 DR P-PSDB; ABB97353.  
 XX An isolated polynucleotide for treating diseases associated with its  
 PT encoded polypeptide such as cancer and multiple sclerosis.  
 XX Claim 1; SEQ ID NO 177; 509pp; English.  
 XX The present invention provides the protein and coding sequences of 444  
 CC novel human proteins. These were isolated from expressed sequences tags  
 CC (ESTs). They can be used to stimulate cell growth, to regulate  
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke  
 CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.  
 CC rheumatoid arthritis, and to treat nervous system disorders e.g.  
 CC Parkinson's disease. The present sequence is a coding sequence of the  
 CC invention  
 XX SQ Sequence 3345 BP; 810 A; 1120 C; 792 G; 623 T; 0 U; 0 Other;  
 Query Match 82.0%; Score 16.4; DB 6; Length 3345;  
 Best Local Similarity 94.4%; Pred. No. 8e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CGAGGCCGCCAGCCCCAC 18  
 DB 274 CGAGGCCGCCAGCCCCAC 291  
 RESULT 17  
 ID ABZ24596 standard; DNA; 19951 BP.  
 XX AC ABZ24596;  
 XX 31-MAR-2003 (first entry)  
 XX Human ovary-specific O1-236 gene.  
 XX Ovary; O1-236; human; nucleoplasmin 2; Npm2; contraceptive;  
 KW antiinfertility; cytostatic; gene therapy; gene; ds.  
 XX OS Homo sapiens.  
 XX Key Location/Qualifiers  
 XX Key 2593..2640  
 XX exon /tag= b  
 XX intron 2641..2769  
 XX /tag= c  
 XX exon 2770..2855  
 XX /tag= d  
 XX intron 2856..2980  
 XX /tag= e  
 XX exon 2981..3106

```
FT intron /*tag= f
FT 3107. .10463
FT /*tag= g
FT 10464. .10557
FT /*tag= h
FT 10558. .11422
FT /*tag= i
FT 11443. .11608
FT /*tag= j
FT 11609. .11877
FT /*tag= k
FT 11833. .11877
FT /*tag= l
FT 11878. .13825
FT /*tag= m
FT 13826. .13859
FT /*tag= n
FT 13860. .13970
FT /*tag= o
FT 13971. .14015
FT /*tag= p
FT 14015
FT CDS /*tag= a
FT /*product= "Human NPM2"
FT /*note= "contains introns"
XX
XX WO200288314-A2.
XX
XX 07-NOV-2002.
XX
XX 26-APR-2002; 2002WO-US013245.
XX
XX 27-APR-2001; 2001US-00844864.
XX (BAY) BAYLOR COLLEGE MEDICINE.
XX (AMHP) WYETH.
XX
XX Matzuk MM, Wang P, Bai Y, Wu X;
XX
XX WPI; 2003-167110/16.
XX P-PSDB; ABP58238.
XX
XX New ovary-specific-genes comprising Ol-180 or Ol-236, useful for
XX decreasing conception or enhancing fertility, or for the preparation of a
XX composition for treating e.g. cancer.
XX
XX Disclosure; Page 119-130; 141pp; English.
XX
XX The present sequence is that of the human ovary-specific Ol-236 gene. Ol-
XX 236 is the mammalian orthologue of Xenopus laevis nucleoplasmin, and the
XX Ol-236 gene has been named Npm2. Experiments in mice have shown that loss
XX of Npm2 results in female infertility and subfertility. The invention
XX provides ovary-specific and oocyte-specific murine and human Ol-180, Ol-
XX 184 and Ol-236 polynucleotides and polypeptides. These genes and their
XX protein products appear to relate to various cell proliferative or
XX degenerative disorders, especially those involving ovarian tumours, such
XX as germ line tumours and granulosa cell tumours, or infertility, such as
XX premature ovarian failure. The invention provides a method for detection
XX of a cell proliferative or degenerative disorder of the ovary, which is
XX associated with the expression of Ol-180, Ol-184 or Ol-236. It also
XX provides a method for treating such disorders by using an agent which
XX suppresses or enhances the respective activities of Ol-180, Ol-184 or Ol-
XX 236, and a method of screening for compounds that interact and/or
XX modulate the expression or activity of the ovary-specific genes. These
XX compounds are possible contraceptive or fertility enhancing agents. The
XX modulator is preferably a polypeptide, small molecule or polynucleotide
XX sequence
XX
XX Sequence 19951 BP; 4619 A; 5201 C; 4927 G; 5204 T; 0 U; 0 Other;
XX
XX Query Match 82.0%; Score 16.4; DB 7; Length 19951;
XX Best Local Similarity 94.4%; Pred. No. 7.9e+02;
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 GAGGCCGCCAGCCCCACA 19
Db 14784 GAGGCCGCCAGCCCCACA 14801

RESULT 18
ABV78124/c
ID ABV78124 standard; DNA; 3417 BP.
XX
XX AC ABV78124;
XX
XX DT 15-NOV-2002 (first entry)
XX
XX DE Human Tiel DNA SEQ ID NO 8.
XX
XX KW RNA inhibition; dsRNA; gene expression inhibitor; oncogene; cytostatic;
XX virucide; protozoacide; gene; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200255693-A2.
XX
XX PD 18-JUL-2002.
XX
XX PF 09-JAN-2002; 2002WO-BP000152.
XX
XX PR 09-JAN-2001; 2001DE-01000586.
XX PR 26-OCT-2001; 2001DE-01055280.
XX PR 29-NOV-2001; 2001DE-01058411.
XX PR 07-DEC-2001; 2001DE-01060151.
XX
XX PA (RIBO-) RIBOPHARMA AG.
XX
XX PI Kreutzer R, Lämmer S, Rost S, Radwiger P;
XX WPI; 2002-590571/63.
XX
XX DR
XX
XX PT Inhibiting expression of target gene, useful e.g. for inhibiting
XX oncogenes, by administering double-stranded RNA complementary to the
XX target and having an overhang.
XX
XX PS Claim 10; Page 116-117; 203pp; German.
XX
XX CC The invention relates to inhibiting expression of a target gene (I) in a
XX cell by introducing an inhibitory RNA (dsRNA1) having a double-stranded
XX structure of at most 49 consecutive bases. At least part of one strand
XX (as1) of dsRNA1 is complementary to (i) and at least one end of dsRNA1
XX has an overhang of 1-4 nucleotides. The method is used to inhibit the
XX expression of a wide range of genes, e.g. oncogenes, cytokine genes etc.
XX in humans, also genes in plasmidium or in viruses or viroids that are
XX pathogenic for humans, animals or plants. Introducing an overhang into
XX dsRNA greatly increases effectiveness for inhibiting gene expression,
XX both in vivo and in vitro and also increases stability and thus the
XX effective concentration inside the cell. The present sequence is that of
XX a gene related to the invention
XX
XX SQ Sequence 3417 BP; 645 A; 1047 C; 1062 G; 663 T; 0 U; 0 Other;
XX
XX Query Match 80.0%; Score 16; DB 6; Length 3417;
XX Best Local Similarity 100.0%; Pred. No. 1.2e+03;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGGCCGCCAGCCCCACA 20
Db 657 GCGGCCGCCAGCCCCACA 642

RESULT 19
ABZ35700/c
ID ABZ35700 standard; DNA; 3417 BP.
XX
XX AC ABZ35700;
```

XX DT 07-FEB-2003 (first entry)  
 XX XX Human Tiel encoding polynucleotide SEQ ID NO 8.  
 XX DE  
 XX KW Double stranded RNA; dsRNA; RNAi; RNA inhibition; cytostatic; virucide;  
 KW Protozoacide; gene expression; antisense; tumour; infection; Plasmodium;  
 KW virus; viroid; anti-GFP; human; HIV; human immunodeficiency virus;  
 KW Hepatitis C virus; human papilloma virus; Gene; ds.  
 XX OS Homo sapiens.  
 XX XX DE10100588-A1.  
 XX PN  
 XX PD 18-JUL-2002.  
 XX PF 09-JAN-2001; 2001DE-01000588.  
 XX PR 09-JAN-2001; 2001DE-01000588.  
 XX PS  
 XX PA (RIBO-) RIBOPHARMA AG.  
 XX PI Kreutzer R, Limmer S, Rost S, Hadwiger P;  
 XX WIPI; 2002-683450/74.  
 XX XX Inhibiting expression of target genes, useful e.g. for treating tumors,  
 PT by introducing into cells two double-stranded RNAs that are complementary  
 PT to the target.  
 XX PS Claim 13; Page 12-13; 100pp; German.  
 XX XX The invention relates to inhibiting expression of a target gene in a cell  
 CC by introducing at least two oligoribonucleotides (dsRNAI and II), both  
 CC with a double-stranded (ds) structure of at most 49 sequential nucleotide  
 CC pairs. At least part of one strand (S1, S2) of the ds structures in each  
 CC of dsRNAI and II are complementary to regions in the target gene. The  
 CC method uses antisense inhibition of gene expression using double stranded  
 CC RNA inhibition (RNAI). The method is particularly used to treat tumours  
 CC or infections, especially by Plasmodium or viruses/viroids (pathogenic on  
 CC humans, animals or plants). The method provides more effective inhibition  
 CC of expression than known methods using a single dsRNA, even at very low  
 CC concentrations. When dsRNA has at least one unpaired nucleotide at the  
 CC end, stability (and thus effective concentration in the cell) is improved  
 CC and efficiency can be increased further by pretreating the cells with  
 CC interferon. The present sequence is that of a target DNA of the invention  
 XX  
 XX SQ Sequence 3417 BP; 645 A; 1047 C; 1062 G; 663 T; 0 U; 0 Other;  
 Query Match 80.0%; Score 16; DB 6; Length 3417;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 5 GCGCCAGCCGCCACAA 20  
 DB 657 GCGCCAGCCGCCACAA 642  
 RESULT 20  
 ID ABX09943/C  
 XX ABX09943 standard; DNA; 3417 BP.  
 XX AC  
 XX AC ABX09943;  
 XX DT 23-JAN-2003 (first entry)  
 XX DE Human Tiel DNA fragment SEQ ID 8.  
 XX DE  
 XX KW Oligoribonucleotide; interferon; oncogene; cytokine; Id; developmental;  
 KW Prion; inhibition; human; ds.  
 XX OS Homo sapiens.  
 XX XX

PN DE10100587-C1.  
 XX 21-NOV-2002.  
 XX PF 09-JAN-2001; 2001DE-01000587.  
 XX PR 09-JAN-2001; 2001DE-01000587.  
 XX PA (RIBO-) RIBOPHARMA AG.  
 XX PI Kreutzer R, Limmer S, Rost S, Hadwiger P;  
 XX WIPI; 2002-742209/81.  
 XX XX Inhibiting expression of target genes, e.g. oncogenes, in cells, by  
 PT introduction of complementary double-stranded oligoribonucleotide, after  
 PT treating the cell with interferon.  
 XX PS Disclosure; Page 17-18; 98pp; German.  
 XX XX This invention describes a novel method for inhibiting expression of a  
 CC target gene by introducing into the cell that contains the target gene at  
 CC least one oligoribonucleotide (dsRNAI) that has a double-stranded (ds)  
 CC structure of not more than 49 consecutive nucleotides (nt), where at  
 CC least a segment of one strand of the ds structure is complementary with  
 CC the target gene and the cells are treated with interferon before  
 CC introduction of dsRNAI. The method is used to inhibit expression of  
 CC target genes, particularly oncogenes, cytokine genes, Id (not defined)  
 CC protein genes; developmental or prion genes, or genes expressed in  
 CC pathogenic organisms (particularly plasmoidal) or in viruses or viroids  
 CC (pathogenic in humans, animals or plants). Treating the cells with  
 CC interferon greatly increases the extent to which dsRNA can inhibit  
 CC expression of the target genes, and the effect is even greater when dsRNA  
 CC are modified to increase their stability. ABX09936-ABX10075 represent  
 CC gene fragments used to illustrate the method of the invention  
 XX  
 XX SQ Sequence 3417 BP; 645 A; 1047 C; 1062 G; 663 T; 0 U; 0 Other;  
 Query Match 80.0%; Score 16; DB 6; Length 3417;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 5 GCGCCAGCCGCCACAA 20  
 DB 657 GCGCCAGCCGCCACAA 642  
 RESULT 21  
 ID ABL91665/C  
 XX ABL91665 standard; DNA; 3417 BP.  
 XX AC  
 XX AC ABL91665;  
 XX DT 28-MAY-2002 (first entry)  
 XX DE Human polynucleotide SEQ ID NO 8.  
 XX DE  
 XX KW Human; HIV; HCV; gene expression; oligoribonucleotide; tumour; pathogen;  
 KW Plasmodium; virus; viroid; cytokine; prion; antisense oligonucleotide;  
 KW cytostatic; virucide; protozoacide; antibacterial; ds.  
 XX OS Homo sapiens.  
 XX XX DE10100586-C1.  
 XX PN  
 XX PD 11-APR-2002.  
 XX PF 09-JAN-2001; 2001DE-01000586.  
 XX PR 09-JAN-2001; 2001DE-01000586.  
 XX PA (RIBO-) RIBOPHARMA AG.  
 XX XX



```

PI Kreutzer R, Limmer S, Rost S, Hadwiger P;
XX WPI; 1993-243152/30.
XX P-PSDB; AAR39820.
XX Nucleic acid encoding tyrosine kinase receptor - for regulating tie
PT activity in diagnosis and treatment of neoplastic diseases involving
PT tumour angiogenesis, wound healing etc.
XX Claim 2; Page 50-53; 73pp; English.
XX The sequence given in AAQ47064 represents the full length 'tie' receptor
CC kinase cDNA. The sequence given in AAQ47065 represents a truncated
CC version of this, wherein nucleotides corresponding to positions 676- 807
CC of the first sequence are absent. Tie is an endothelial cell receptor
CC tyrosine kinase and contains immunoglobulin- (Ig-) and epidermal growth
CC factor- (EGF-) like repeats. The tie gene is expressed in the endothelial
CC cells of blood vessels. Tie levels may indicate onset/progression of
CC diseases involving endothelial cells and their tie receptors, such as
CC neoplastic diseases involving tumour angiogenesis, wound healing,
CC thromboembolic diseases, atherosclerosis and inflammatory diseases.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 3845 BP; 743 A; 1179 C; 1155 G; 768 T; 0 U; 0 Other;
Query Match 80.0%; Score 16; DB 2; Length 3845;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GCGCCAGCCCCACAA 20
Db 693 GCGCCAGCCCCACAA 678

RESULT 23
ABZ34892/C
ID ABZ34892 standard; cDNA; 3845 BP.
XX AC ABZ34892;
XX AC ABZ34892;
DT 05-FEB-2003 (first entry)
XX Human gene expression profile polynucleotide SEQ ID NO 4.
XX Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;
XX bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;
XX tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;
XX gene expression; gene; ss.
XX Homo sapiens.
XX OS Homo sapiens.
XX PN WO200274979-A2.
XX PD 26-SEP-2002.
XX PF 20-MAR-2002; 2002WO-US008456.
XX PR 20-MAR-2001; 2001US-0276947P.
XX PA (ORTH ) ORTHO CLINICAL DIAGNOSTICS INC.
XX PI Wan J, Wang Y;
XX WPI; 2002-740862/80.
XX New gene expression profile generated from primary, endothelial,
PT epithelial, and muscle cell types, useful for identifying disease
PT pathologies involving alterations of gene expression, e.g. cancer.
XX Claim 1; Page 223-224; 850pp; English.
XX The invention relates to a gene expression profile comprising one or more
CC genes (ABZ34899-ABZ35692) and generated from a cell type. The cell type
CC is a coronary artery endothelium, umbilical artery or vein endothelium,
CC aortic endothelium, dermal microvascular endothelium, pulmonary artery

```



PS Example 2; Page 232-233; 291pp; English.

XX The present invention relates to methods and compositions for detecting  
CC an angiogenesis-associated transcript in a cell in a patient. The method  
CC involves contacting a biological sample from the patient with a  
CC polynucleotide that selectively hybridizes to a sequence at least 80%  
CC identical to any of the angiogenesis-associated human polynucleotide  
CC sequences given in the specification. These angiogenesis-associated  
CC polynucleotide sequences comprise genes that exhibit changes in  
CC expression levels as a function of time in tissue undergoing  
CC angiogenesis. The method and the polynucleotide sequences of the  
CC invention are useful for diagnosing and treating angiogenesis and  
CC angiogenesis-associated diseases e.g. cancer. The polynucleotide  
CC sequences are also useful in the gene therapy of such disorders. The  
CC angiogenesis-associated proteins encoded by the polynucleotide sequences  
CC are useful as a vaccine for therapeutic and prophylactic immunisation.  
CC ABX08739-ABX08853 represent angiogenesis-associated polynucleotide  
XX sequences

SQ Sequence 3845 BP; 743 A; 1179 C; 1155 G; 768 T; 0 U; 0 Other;

Query Match 80.0%; Score 16; DB 7; Length 3845;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 GCGCCCGAGCCCAAA 20  
| | | | | | | | | |  
Db 693 GCGCCCGAGCCCAAA 678

RESULT 26

ABX12538/c  
ID ABX12538 standard; cDNA; 3845 BP.

AC ABX12538;

DT 10-MAY-2003 (first entry)

DE cDNA encoding human Tie receptor tyrosine kinase 1 (Tie 1).

XX Cytostatic; vasodilator; antiinflammatory; cardiant; gene therapy;  
KW ligand-receptor binding modulator; ephrin ligand; angiogenesis;  
KW lymphangiogenesis; aberrant Ephrin-Tie biology; cell growth disorder;  
KW cell migration disorder; cell proliferation disorder; neovascularisation;  
KW ischaemia; infarction; tissue graft; transplant; human; tie 1;  
KW tie receptor tyrosine kinase 1; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
FH 37..3453  
CDS /tag= a

FT /product= "Tie 1"  
FT /note= "Tie receptor tyrosine kinase"

XX WO2003004529-A2.

XX 16-JAN-2003.

XX 02-JUL-2002; 2002WO-IB002524.

XX 02-JUL-2001; 2001US-0302960P.

XX (LICN) LICENTIA LTD.

XX Alitalo K, Kubo H;

XX WPI; 2003-210341/20.

XX P-PSDB; ABU07837.

XX Identifying modulators of binding between a Tie receptor tyrosine kinase  
PT and an Ephrin ligand, useful for promoting neovascularization, comprises  
PT contacting a Tie receptor with an Ephrin in the presence of a putative

PT modulator.

XX Disclosure; Page 56-61; 199pp; English.

XX The invention describes a method of identifying a modulator of binding  
CC between a Tie receptor tyrosine kinase and an Ephrin ligand. The method  
CC comprises contacting a Tie receptor composition with an Ephrin  
CC composition in the presence and in the absence of a putative modulator  
CC compound, and detecting the binding between Tie receptor and the Ephrin  
CC in the presence and in the absence of the putative modulator. The method  
CC is useful for identifying a modulator of binding between a Tie receptor  
CC tyrosine kinase and an Ephrin ligand. Modulators identified from the  
CC method are useful in modulating angiogenic processes, including  
CC lymphangiogenesis, for treating diseases associated with aberrant Ephrin-  
CC Tie biology, aberrant growth, migration or proliferation of cells that  
CC express a Tie receptor, or for promoting growth of vessel or  
CC neovascularisation (e.g. ischaemic tissue, an infarction, a new or  
CC chronic compound, or a tissue graft or transplant). This sequence encodes  
XX human Tie receptor tyrosine kinase 1 (Tie 1)

SQ Sequence 3845 BP; 743 A; 1179 C; 1155 G; 768 T; 0 U; 0 Other;

Query Match 80.0%; Score 16; DB 7; Length 3845;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 GCGCCCGAGCCCAAA 20  
| | | | | | | | | |  
Db 693 GCGCCCGAGCCCAAA 678

RESULT 27

ABX63148/c  
ID ABX63148 standard; cDNA; 3914 BP.

AC ABX63148;

DT 25-FEB-2003 (first entry)

XX Human cDNA #148 differentially expressed in activated vascular tissue.

XX Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;  
KW hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;  
KW gene therapy; vascular disease; cancer; coronary; artery disease;  
KW hypertension; diabetes; pre-eclampsia; restenosis;  
KW ischaemia-reperfusion injury; stroke.

XX Homo sapiens.

XX US2002137081-A1.

XX 26-SEP-2002.

XX 08-JAN-2002; 2002US-00044090.

XX 28-JUL-2000; 2000US-0222469P.

XX 08-JAN-2001; 2001US-0260483P.

XX (BAND/) BANDMAN O.

XX Bandman O;

XX WPI; 2003-110597/10.

XX Combination for diagnosing, staging, treating, or monitoring the  
PT progression of treatment of a vascular disease, e.g. atherosclerosis,  
PT comprises several cDNAs that are differentially expressed in activated  
PT vascular tissue.

XX Claim 1; Page; 18pp; English.

XX This invention relates to a combination comprising several cDNAs that are  
CC differentially expressed in activated vascular tissue. The invention also

CC discloses a high throughput method for detecting differentially expressed  
 CC cDNAs in a sample. The cDNAs of the invention may have  
 CC antiarteriosclerotic; cytotatic; cardiant; hypotensive; antidiabetic;  
 CC gynaecological; vasotropic and cerebroprotective activities and may be  
 CC used in gene therapy. The cDNAs of the invention may be used in a high-  
 CC throughput methods for detecting differential expression of one or more  
 CC cDNAs in a sample, or screening several molecules or compounds to  
 CC identify a molecule or compound that specifically binds a cDNA of the  
 CC invention. A protein encoded by the cDNA may be used to screen several  
 CC molecules or compounds to identify a ligand that specifically binds to  
 CC the protein, or to produce or purify an antibody to the protein that can  
 CC be used to detect a protein in a sample or purify a natural or  
 CC recombinant protein from a sample. The nucleotides may be useful for  
 CC diagnosing, staging, treating, or monitoring the progression of treatment  
 CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery  
 CC disease, hypertension, diabetes, pre-eclampsia, ischaemia- reperfusion  
 CC injury, restenosis, or stroke. The cDNAs can also be used for large-scale  
 CC genetic or gene expression analysis of several new nucleic acid  
 CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for  
 CC diagnosing pre-pathologic disorders, and chronic or acute diseases  
 CC associated with abnormalities in the expression, amount or distribution  
 CC of the protein. The present sequence represents a cDNA of the invention  
 CC that is differentially expressed in activated vascular tissue. Note: the  
 CC sequence data for this patent did not form part of the specification, but  
 CC was obtained in electronic format directly from USPTO at  
 CC <http://seqdata.uspto.gov/sequence.html?docID=20020137081>  
 XX  
 XX Sequence 3914 BP; 768 A; 1205 C; 1168 G; 773 T; 0 U; 0 Other;

Query Match 80.0%; Score 16; DB 7; Length 3914;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGCCAGCCCCACAA 20  
 DB 754 GCGCCAGCCCCACAA 739

RESULT 28  
 ACH03957/c  
 ID ACH03957 standard; cDNA; 3914 BP.  
 XX  
 XX ACH03957;  
 XX  
 XX 26-SEP-2003 (first entry)  
 XX Human cDNA differentially expressed in lung cancer #162.  
 XX Gene therapy; emphysema; ss; gene; chronic obstructive pulmonary disease;  
 XX respiratory disorder; lung cancer; asthma; human.  
 XX Homo sapiens.  
 XX US2003065157-A1.  
 XX  
 XX 03-APR-2003.  
 XX  
 XX 04-APR-2002; 2002US-00116802.  
 XX  
 XX 04-APR-2001; 2001US-0281593P.  
 XX (LASE/) LASEK A W.  
 XX Lasek AW;  
 XX WPI; 2003-540803/51.  
 XX New combination comprising cDNAs that are differentially expressed in  
 XX respiratory disorders, useful for diagnosing or treating respiratory  
 XX disorders e.g., lung cancer, chronic obstructive pulmonary disease,  
 XX emphysema or asthma.  
 XX Claim 1; Page; 39pp; English.

XX The invention relates to a combination comprising cDNAs or their  
 CC complements that are differentially expressed in respiratory disorder.  
 CC The combination is useful for preparing a composition for diagnosing or  
 CC treating respiratory disorders e.g. lung cancer, chronic obstructive  
 CC pulmonary disease, emphysema or asthma. The present sequence represents  
 CC human cDNA differentially expressed during lung cancer  
 XX  
 XX Sequence 3914 BP; 768 A; 1205 C; 1168 G; 773 T; 0 U; 0 Other;  
 SQ  
 Query Match 80.0%; Score 16; DB 8; Length 3914;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGCCAGCCCCACAA 20  
 DB 754 GCGCCAGCCCCACAA 739

RESULT 29  
 AAZ32059/c  
 ID AAZ32059 standard; DNA; 8670 BP.

XX  
 AC AAZ32059;

XX 10-JAN-2000 (first entry)

XX Human METH2 related EST X16619.

XX Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;  
 KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;  
 KW angiogenesis inhibitor; abnormal wound healing; inflammation;  
 KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;  
 KW diabetic retinopathy; macula degeneration; haemangioma; detection;  
 KW arterial-venous malformation; immune deficiency; ss.

XX Homo sapiens.

XX WO9937660-A1.

XX 29-JUL-1999.

XX 22-JAN-1999; 99WO-US001313.

XX 23-JAN-1998; 98US-0072988P.

XX 28-AUG-1998; 98US-0098539P.

XX (TRUE/) IRUELA-ARISPE L.

XX (HAST/) HASTINGS G A.

XX (RUBE/) RUBEN S M.

XX Iruela-Arispe L, Hastings GA, Ruben SM;

XX WPI; 1999-590684/50.

XX New isolated metalloprotease thrombospondin polypeptides, useful for  
 XX treating hyperproliferative disorders, cancers or autoimmune disorders.

XX Disclosure; Page 431-437; 457pp; English.

XX AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human  
 CC metalloprotease thrombospondin (METH) proteins METH1 and METH2  
 CC respectively. METH1 and METH2 have been found to be potent inhibitors of  
 CC angiogenesis both in vitro and in vivo. They can be used for treating  
 CC cancer and other disorders related to angiogenesis including abnormal  
 CC wound healing, inflammation, rheumatoid arthritis, psoriasis, endometrial  
 CC bleeding disorders, diabetic retinopathy, some forms of macula  
 CC degeneration, haemangiomas, and arterial-venous malformations. They may  
 CC be useful in treating deficiencies or disorders of the immune system, by  
 CC activating or inhibiting the proliferation, differentiation, or  
 CC mobilisation (chemotaxis) of immune cells. The etiology of these immune  
 CC deficiencies or disorders may be genetic, somatic, such as cancer or some  
 CC autoimmune disorders, acquired (e.g. by chemotherapy or toxins), or

PS  
XX  
Claim 14: Page 731-737: 768pp: English.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 934.603 Seconds  
(without alignments)  
639.034 Million cell updates/sec

Title: US-10-624-714-11

Perfect score: 20

Sequence: 1 cgaggcgccagcccccacaa 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

EST:\*

1: em\_estba:\*\*

2: em\_esthum:\*\*

3: em\_estin:\*\*

4: em\_estmu:\*\*

5: em\_estov:\*\*

6: em\_estpl:\*\*

7: em\_estro:\*\*

8: em\_htc:\*\*

9: gb\_est1:\*\*

10: gb\_est2:\*\*

11: gb\_htc:\*\*

12: gb\_est3:\*\*

13: gb\_est4:\*\*

14: gb\_est5:\*\*

15: em\_estfun:\*\*

16: em\_estom:\*\*

17: em\_gss\_hum:\*\*

18: em\_gss\_inv:\*\*

19: em\_gss\_pln:\*\*

20: em\_gss\_vrt:\*\*

21: em\_gss\_fun:\*\*

22: em\_gss\_mam:\*\*

23: em\_gss\_mus:\*\*

24: em\_gss\_pro:\*\*

25: em\_gss\_rod:\*\*

26: em\_gss\_phg:\*\*

27: em\_gss\_vrl:\*\*

28: gb\_gss1:\*\*

29: gb\_gss2:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18.4	92.0	541	9	AU192152
C 2	17.4	87.0	144	9	AU192152 AU192152
C 3	17.4	87.0	183	12	AZ207487 mv9906.r
C 4	17.4	87.0	221	10	BI298853 UI-R-CV2-BF333695 QV0-CS000

78	16.4	82.0	395	12	BG896292	BG896292 HOA29-1-A	151	16	80.0	843	29	CC472932	CC472932	CH240 297
79	16.4	82.0	398	10	AW161847	AW161847 au88c02.y	C 152	16	80.0	850	10	BF979359	BF979359 602288449	
80	16.4	82.0	441	14	W39461	W39461 xc229001.r1	C 153	16	80.0	858	9	AL553654	AL553654 602288449	
81	16.4	82.0	448	13	BQ842457	BQ842457 WHE2993.C	C 154	16	80.0	857	13	BQ893725	BQ893725 AGENCOURT	
82	16.4	82.0	456	9	AA422277	AA422277 vf44c12.f	C 155	16	80.0	857	9	AL574074	AL574074 603631824	
83	16.4	82.0	459	29	CE350973	CE350973 tigr-gss-	C 156	16	80.0	1100	9	AL574074	AL574074 603631824	
84	16.4	82.0	466	12	BM402175	BM402175 SLA001B04	C 157	16	80.0	1164	12	BM010611	BM010611 603631824	
85	16.4	82.0	473	12	BM402990	BM402990 K-EST0120	C 158	16	80.0	1201	9	AL534990	AL534990 603631824	
86	16.4	82.0	473	28	AL191918	AL191918 RPI11-45	C 159	16	80.0	1201	9	AL547090	AL547090 603631824	
87	16.4	82.0	486	9	AI179353	AI179353 EST223050	C 160	16	80.0	1201	13	BA443821	BA443821 603631824	
88	16.4	82.0	487	14	CB431452	CB431452 607429 MA	C 161	16	80.0	1227	28	B2579507	B2579507 msh2 6387	
89	16.4	82.0	491	14	CD528841	CD528841 Leuk0N3.5	C 162	16	80.0	2356	11	BC039336	BC039336 Homc_sapi	
90	16.4	82.0	501	10	BF601870	BF601870 268964 MA	C 163	16	80.0	3264	29	AY411079	AY411079 Pan trogl	
91	16.4	82.0	511	29	CNS0330H	AL226394 Tetraodon	C 164	15.8	79.0	3359	29	AY411078	AY411078 Homo sapi	
92	16.4	82.0	517	12	BM402310	BM402310 SLA002G01	C 165	15.8	79.0	167	28	AZ266956	AZ266956 RPI1-23-4	
93	16.4	82.0	528	12	EG707748	EG707748 602671275	C 166	15.8	79.0	197	28	AZ266956	AZ266956 RPI1-23-3	
94	16.4	82.0	528	13	BQ807985	BQ807985 NISC.kk12	C 167	15.8	79.0	233	29	CG606693	CG606693 OST284846	
95	16.4	82.0	530	14	CH155654	CH155654 K-EST0214	C 168	15.8	79.0	247	14	W05051	W05051 za78G06.r1	
96	16.4	82.0	530	28	AQ450862	AQ450862 HS 5173.A	C 169	15.8	79.0	248	12	BI035085	BI035085 BM2-NN016	
97	16.4	82.0	542	9	AI929330	AI929330 au5b06.y	C 170	15.8	79.0	259	10	AM406781	AM406781 UI-HF-BL0	
98	16.4	82.0	547	9	AI231633	AI231633 EST228321	C 171	15.8	79.0	262	9	AA768688	AA768688 ca68H09.s	
99	16.4	82.0	547	9	AI231634	AI231634 EST228322	C 172	15.8	79.0	268	14	CF380403	CF380403 lac46g01.	
100	16.4	82.0	605	12	BZ271476	BZ271476 BU271476	C 173	15.8	79.0	269	12	BI289215	BI289215 UI-R-DK0-	
101	16.4	82.0	610	9	AI231665	AI231665 EST228353	C 174	15.8	79.0	275	29	CG536757	CG536757 OST125046	
102	16.4	82.0	614	14	CH156703	CH156703 EST228323	C 175	15.8	79.0	287	10	BM402788	BM402788 BU42788	
103	16.4	82.0	622	9	AI231635	AI231635 EST228323	C 176	15.8	79.0	291	10	BM402788	BM402788 BU42788	
104	16.4	82.0	651	12	BG713097	BG713097 602699027	C 177	15.8	79.0	295	12	BM707654	BM707654 UI-E-C11-	
105	16.4	82.0	651	28	BZ841816	BZ841816 CH240 193	C 178	15.8	79.0	295	14	CB149519	CB149519 K-EST0206	
106	16.4	82.0	658	14	CH044600	CH044600 NISC.GC05	C 179	15.8	79.0	298	9	AI614455	AI614455 VG79G04.Y	
107	16.4	82.0	664	9	AI517534	AI517534 to27C01.x	C 180	15.8	79.0	298	10	B489359	B489359 BA489359	
108	16.4	82.0	665	14	CH153762	CH153762 K-EST0211	C 181	15.8	79.0	300	9	AU099612	AU099612 AU099612	
109	16.4	82.0	670	12	BG773390	BG773390 602722179	C 182	15.8	79.0	301	10	BE314097	BE314097 601146436	
110	16.4	82.0	679	14	CD873830	CD873830 AZ03.100M	C 183	15.8	79.0	301	13	BY362043	BY362043 BY362043	
111	16.4	82.0	710	29	CG819701	CG819701 100006C05	C 184	15.8	79.0	315	9	AA312493	AA312493 EST183161	
112	16.4	82.0	713	14	CD373798	CD373798 WHE2603.G	C 185	15.8	79.0	318	13	BY362148	BY362148 BY362148	
113	16.4	82.0	715	12	BJ276704	BJ276704 BZ276704	C 186	15.8	79.0	325	12	BM002864	BM002864 103110780	
114	16.4	82.0	760	12	B4523948	B4523948 603081119	C 187	15.8	79.0	325	10	BE394018	BE394018 601312373	
115	16.4	82.0	766	10	BF687885	BF687885 602066957	C 188	15.8	79.0	325	13	BY362056	BY362056 BY362056	
116	16.4	82.0	772	10	BE314620	BE314620 601154377	C 189	15.8	79.0	329	29	CE579378	CE579378 tigr-gss-	
117	16.4	82.0	786	12	BG773311	BG773311 602722079	C 190	15.8	79.0	330	12	BM526779	BM526779 sal45f05.	
118	16.4	82.0	834	10	BF690786	BF690786 602246766	C 191	15.8	79.0	349	14	TS1154	TS1154 YP94e04.r1	
119	16.4	82.0	834	10	B5973481	B5973481 601652317	C 192	15.8	79.0	353	9	AA412608	AA412608 t955d12.r	
120	16.4	82.0	865	29	CG680296	CG680296 OGVAR17TV	C 193	15.8	79.0	353	9	AA346011	AA346011 EST52101	
121	16.4	82.0	896	13	BQ489596	BQ489596 AGENCOURT	C 194	15.8	79.0	355	10	BP474739	BP474739 RC6-EN008	
122	16.4	82.0	904	13	BQ672920	BQ672920 AGENCOURT	C 195	15.8	79.0	355	12	BP430072	BP430072 BP430072	
123	16.4	82.0	971	29	CG602917	CG602917 OGURH59TV	C 196	15.8	79.0	359	12	BP428605	BP428605 BP428605	
124	16.4	82.0	1014	10	BE373067	BE373067 601236682	C 197	15.8	79.0	360	9	AJ465590	AJ465590 AJ465590	
125	16.4	82.0	1086	28	CG240428	CG240428 CH261-102	C 198	15.8	79.0	363	12	BP430674	BP430674 BP430674	
126	16.4	82.0	1105	12	BM921048	BM921048 AGENCOURT	C 199	15.8	79.0	364	12	BP429883	BP429883 BP429883	
127	16.4	82.0	1201	9	AL568877	AL568877 AL568877	C 200	15.8	79.0	367	9	AA268348	AA268348 va41e11.r	
128	16.4	82.0	1201	9	AG023045	AG023045 Pan trogl	C 201	15.8	79.0	371	10	AW196322	AW196322 xm31g12.x	
129	16.4	82.0	1210	11	BC039110	BC039110 Homc_sapi	C 202	15.8	79.0	371	12	BP428751	BP428751 BP428751	
130	16.4	82.0	1299	10	BF850180	BF850180 IL5-EN008	C 203	15.8	79.0	372	10	AW406159	AW406159 UI-HF-BL0	
131	16.4	82.0	301	10	BF847831	BF847831 IL5-EN008	C 204	15.8	79.0	383	9	AA345937	AA345937 EST52042	
132	16.4	82.0	301	10	BF849488	BF849488 IL5-EN008	C 205	15.8	79.0	383	9	AA417709	AA417709 t955d12.r	
133	16.4	82.0	302	10	BF849487	BF849487 IL5-EN008	C 206	15.8	79.0	383	14	H63635	H63635 YP34B10.r1	
134	16.4	82.0	303	10	BF850172	BF850172 IL5-EN008	C 207	15.8	79.0	383	14	H77364	H77364 YU12e10.r1	
135	16.4	82.0	318	10	BF847825	BF847825 IL5-EN008	C 208	15.8	79.0	384	29	CG579414	CG579414 OST717709	
136	16.4	82.0	324	10	BF847824	BF847824 IL5-EN008	C 209	15.8	79.0	384	9	AA305775	AA305775 EST176774	
137	16.4	82.0	337	10	BF847819	BF847819 IL5-EN008	C 210	15.8	79.0	384	9	AA305775	AA305775 EST176774	
138	16.4	82.0	337	10	BF847821	BF847821 IL5-EN008	C 211	15.8	79.0	384	14	T72728	T72728 YU12e10.r1	
139	16.4	82.0	337	10	BF847959	BF847959 IL5-EN008	C 212	15.8	79.0	393	10	EG074341	EG074341 H3133H10-	
140	16.4	82.0	350	13	BY157875	BY157875 BY157875	C 213	15.8	79.0	393	9	AW057912	AW057912 W35905.x	
141	16.4	82.0	352	10	BF850044	BF850044 IL5-EN008	C 214	15.8	79.0	398	13	BQ304658	BQ304658 CM4-B085	
142	16.4	82.0	353	10	BF850211	BF850211 IL5-EN008	C 215	15.8	79.0	398	12	BM923459	BM923459 AGENCOURT	
143	16.4	82.0	361	14	H33054	H33054 YG83B08.s1	C 216	15.8	79.0	398	14	T10017	T10017 seq1395 b4H	
144	16.4	82.0	563	29	CE549613	CE549613 tigr-gss-	C 217	15.8	79.0	404	12	BP428601	BP428601 BP428601	
145	16.4	82.0	595	12	BG895953	BG895953 359665 MA	C 218	15.8	79.0	407	14	TS5830	TS5830 EST92099 Hu	
146	16.4	82.0	596	12	B1066398	B1066398 pgfin.pk0	C 219	15.8	79.0	408	13	BY217667	BY217667 BY217667	
147	16.4	82.0	686	10	B3641987	B3641987 BB641987	C 220	15.8	79.0	413	10	AW577530	AW577530 EST389639	
148	16.4	82.0	692	29	AG147725	AG147725 Pan trogl	C 221	15.8	79.0	420	10	AW674963	AW674963 ba60d01.y	
149	16.4	82.0	783	12	BI463932	BI463932 603202673	C 222	15.8	79.0	425	9	AA943419	AA943419 EST198918	
150	16.4	82.0	808	13	BU503249	BU503249 AGENCOURT	C 223	15.8	79.0	427	10	AW913071	AW913071 uf48f05.y	



```
224 15.8 79.0 430 9 AI802084
225 15.8 79.0 431 12 BM840118 K-EST0117
226 15.8 79.0 432 28 AZ039431
227 15.8 79.0 433 14 H65577
228 15.8 79.0 434 10 AW414902
229 15.8 79.0 435 14 BE674158
230 15.8 79.0 436 14 W78068
231 15.8 79.0 437 10 BE468248
232 15.8 79.0 438 14 W39608
233 15.8 79.0 439 14 PR0038212
234 15.8 79.0 440 29 CF555178
235 15.8 79.0 441 28 BZ608882
236 15.8 79.0 442 14 CB785676
237 15.8 79.0 443 9 AI087321
238 15.8 79.0 444 14 H70038
239 15.8 79.0 445 9 AI364711
240 15.8 79.0 446 12 BI748915
241 15.8 79.0 447 9 AI310009
242 15.8 79.0 448 29 CE762574
243 15.8 79.0 449 9 AI236897
244 15.8 79.0 450 28 AZ589021
245 15.8 79.0 451 13 BX519576
246 15.8 79.0 452 13 BQ143077
247 15.8 79.0 453 14 CB142053
248 15.8 79.0 454 28 AZ748330
249 15.8 79.0 455 10 BE09873
250 15.8 79.0 456 28 AQ339508
251 15.8 79.0 457 10 BB759368
252 15.8 79.0 458 14 CD545362
253 15.8 79.0 459 9 AI160577
254 15.8 79.0 460 12 BM687393
255 15.8 79.0 461 14 CD538335
256 15.8 79.0 462 9 AI433386
257 15.8 79.0 463 10 AW553832
258 15.8 79.0 464 9 AI225675
259 15.8 79.0 465 10 AW671976
260 15.8 79.0 466 9 AI154624
261 15.8 79.0 467 9 AI962454
262 15.8 79.0 468 12 BQ973671
263 15.8 79.0 469 10 BQ017350
264 15.8 79.0 470 28 AZ200586
265 15.8 79.0 471 9 AI047721
266 15.8 79.0 472 13 BX511881
267 15.8 79.0 473 10 BE878880
268 15.8 79.0 474 9 AI306825
269 15.8 79.0 475 12 BM701338
270 15.8 79.0 476 10 AW674953
271 15.8 79.0 477 10 BE232321
272 15.8 79.0 478 10 AW953545
273 15.8 79.0 479 10 BF1313796
274 15.8 79.0 480 13 BQ189194
275 15.8 79.0 481 10 AW139551
276 15.8 79.0 482 14 CD542765
277 15.8 79.0 483 13 BQ830017
278 15.8 79.0 484 14 CB158908
279 15.8 79.0 485 14 CD542751
280 15.8 79.0 486 12 BQ614138
281 15.8 79.0 487 10 CF534607
282 15.8 79.0 488 29 CE156722
283 15.8 79.0 489 9 AI131161
284 15.8 79.0 490 9 AI154675
285 15.8 79.0 491 14 CD538158
286 15.8 79.0 492 10 BE312164
287 15.8 79.0 493 29 CE069728
288 15.8 79.0 494 13 BX668561
289 15.8 79.0 495 12 BM352370
290 15.8 79.0 496 14 CB112642
291 15.8 79.0 497 9 AI673892
292 15.8 79.0 498 10 BE871180
293 15.8 79.0 499 10 BE621823
294 15.8 79.0 500 10 AW272400
295 15.8 79.0 501 9 AI043584
296 15.8 79.0 502 10 AW248200
```

```
AI802084 tx24f02.x
BM840118 K-EST0117
AZ039431 RPI-23-3
H65577 yr62a09.r1
AW414902 48905.MAR
BE674158 7d76a05.x
W78068 zd72e07.r1
BE468248 hz71b02.x
W39608 zc19e04.r1
AI125713 Fugu rubr
CF555178 l115006A1
BZ608882 WHAAB49TF
CB785676 AMGNUN9.C
AI087321 oz45f09.x
H70038 yu73b11.r1
AI364711 qz07b05.x
BI748915 rc83f11.y
AI310009 EST180922
CE762574 tigr-gss-
AA236897 zs4309.s
AZ589021 LM0387P07
BX519576 BX519576
BQ143077 fmb1c.pk0
CB142053 K-EST0195
AZ748330 CQ04802.f
BE09873 RC5-BT070
AQ339508 HS 5022.A
BB759368 BF759368
CD545362 BQ266C01-
AI160577 qc83c04.x
BM687393 UI-E-CQ1-
CD538335 BQ202C11-
AI433386 tgr3a06.x
AW553832 L0232D01-
AI225675 uj14b12.y
AW671976 LGI 353.D
AI154624 AU154624-
AI962454 wq52e06.x
BQ973671 601680849
BQ017350 UI-H-D11-
AZ200586 SP 1039.A
AI047721 DKEZp5860
BX511881 BX511881
BE878880 601493175
AI306825 EST177820
BM701338 UI-E-EJ0-
AW674953 Bq60c02.y
BE232321 137313.NA
AW953545 EST365615
BF1313796 601900471
BQ189194 UI-E-EJ1-
AW139551 UI-H-B11-
CD542765 BQ243E07-
BQ830017 Lf61n2041
CB158908 K-EST0218
CD542751 BQ243D10-
BQ614138 BQ061438
CF534607 UI-M-G10-
CE156722 tigr-gss-
AI131161 qc15c01.x
AI154675 AU154675
CD538158 BQ200G10-
BE312164 601152459
CE069728 tigr-gss-
BX668561 BX668561
BM352370 i98907.y
CB112642 K-EST0154
AI673892 AL673892
BE871180 601448752
BE621823 601494075
AW272400 xv14b03.x
AI043584 DKFZp434F
AW248200 2819721.5
```

297 15.8 79.0 592 12 BI796574  
298 15.8 79.0 593 14 CB269034  
299 15.8 79.0 593 29 CE442639  
300 15.8 79.0 596 10 AW553048

ALIGNMENTS

RESULT 1  
AI192152/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

AI192152 Porphyra yezoensis TU-1 sporophyte  
CDNA clone PFL026f01\_r 5', mRNA sequence.  
AI192152  
AI192152.1 GI:31930508  
EST.  
Porphyra yezoensis  
Porphyra yezoensis  
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;  
Porphyra.  
1 (bases 1 to 541)  
Asanizu,E., Nakajima,M., Kitade,Y., Saga,N., Nakamura,Y. and  
Tabata,S.  
COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS  
OF PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG  
FREQUENCY ANALYSIS  
J. Phycol. 39 (5), 923-930 (2003)  
Contact: Erika Asanizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asanizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

1. 541  
/organism="Porphyra yezoensis"  
/mol\_type="mRNA"  
/strain="TU-1"  
/db\_xref="taxon:2788"  
/clone="PFL026f01.x"  
/dev\_stage="sporophytes"  
/clone\_lib="Porphyra yezoensis TU-1 sporophytes"

ORIGIN  
Query Match 92.0%; Score 18.4; DB 9; Length 541;  
Best Local Similarity 95.0%; Pred. No. 3 5e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCAACAA 20  
|||||  
Db 156 CGAGGCGCCAGCCCAACAA 137

RESULT 2  
AA207487/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

AA207487  
mv89g06.r1 GuayWoodford Beier mouse kidney day 7 Mus musculus cdna  
clone IMAGE:662266 5', mRNA sequence.  
AA207487  
AA207487.1 GI:1804059  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 144)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)

## COMMENT

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:408114

Seq primer: -28mb3 rev1 ET from Amersham  
High quality sequence stop: 133.

## FEATURES

## source

1. .144  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:662266"  
/tissue\_type="kidney"  
/dev\_stage="juvenile (7 days old)"  
/lab\_host="SOLR (kanamycin resistant)"  
/clone\_lib="GuayWoodford Beier mouse kidney day 7"  
/note="Organ: kidney; Vector: pBluescript SK-; Site 1:  
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector;  
-5' adaptor sequence: 5' GAATTCGACGACGAG 3' -3' adaptor  
sequence: 5' CTCGAGTTTCTTTTCTTTT 3' Library provided  
by Lisa Guay-Woodford."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 144;  
Best Local Similarity 94.7%; Pred. No. 6.6e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCAGCCACCA 19  
Db 108 CGAGGACCCAGCCACCA 90

## RESULT 3

BI298853/c  
LOCUS 183 bp mRNA linear EST 20-JUL-2001  
DEFINITION UI-R-CV2-chv-d-06-0-UI-s1 UI-R-CV2 Rattus norvegicus cDNA clone  
UI-R-CV2-chv-d-06-0-UI 3', mRNA sequence.

ACCESSION BI298853  
VERSION BI298853.1 GI:14975133

KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 183)  
AUTHORS Bonaldo,M.P., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

PUBMED 8889548

COMMENT Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565

Email: bento-soares@uiowa.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to verify it as a clone from the  
normalized rat eye library cDNA library Preparation: M.B. Soares

## FEATURES

## source

1. .183  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-CV2-chv-d-06-0-UI"  
/dev\_stage="ADULT"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-CV2"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; CV2 is a  
subtracted library derived from a normalized eye library  
(rREE2), constructed according to the procedure described  
by Bonaldo, Lennon & Soares (Genome Research Genome 6:  
791-806, 1996). For construction of the CV2 library,  
plasmid DNA from the normalized rat eye library was  
electroporated into competent bacteria for production of  
single-stranded circular DNA. This was then used as a  
tracer in a subtractive hybridization with a driver  
comprising: a) a set of about 1,000 arrayed clones from  
the non-normalized eye library CV0 (PCR amplified inserts  
from a plasmid DNA template preparation representing  
plates R-CV0-BRH through R-CV0-BRR). This represented 20%  
of the final driver population. b) A set of about 4,000  
arrayed clones from the normalized eye library CV1 (PCR  
amplified inserts from a plasmid DNA template preparation  
representing plates R-CV1-BRS through R-CV1-BSC, R-CV1-BSE  
through R-CV1-BTC, and R-CV1-BV0 through R-CV1-BVU). This  
represented 80% of the final driver population.

TAG TISSUE=rat eye

TAG\_LIB=UI-R-CV2

TAG\_SEQ=CAGCC

## ORIGIN

Query Match 87.0%; Score 17.4; DB 12; Length 183;  
Best Local Similarity 94.7%; Pred. No. 6.9e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCAGCCACCA 20

Db 82 GGGGCGCCAGCCACCA 64

## RESULT 4

BF333695

LOCUS 221 bp mRNA linear EST 22-NOV-2000

DEFINITION QV0-CS0009-220900-401-h07 CS0009 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF333695

VERSION BF333695.1 GI:11304443

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 221)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,

O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.U.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PUBMED 10737800

COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics

Lab Clone distribution: clones will be available through Research  
Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=Yes.

Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimposoludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV0&t2=QV0-CS0009-  
220900-401-h07&t3=2000-09-22&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 15  
High quality sequence stop: 221.

## FEATURES

source

1. 221  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="adult"  
/clone\_lib="CS0009"  
note="Organ: colon\_est; Vector: puc18; Site 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 221;  
Best Local Similarity 94.7%; Pred. No. 7.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGGCCCCAGCCCAAA 20

DB 107 GAGGGCCCCAGCCCAAA 125

## RESULT 5

AI713710/c

LOCUS AI713710 222 bp mRNA linear EST 08-JUN-1999  
DEFINITION UI-R-AGI-aao-b-11-0-UI.s1 UI-R-AGI Rattus norvegicus cDNA clone  
UI-R-AGI-aao-b-11-0-UI 3', mRNA sequence.

ACCESSION AI713710

VERSION AI713710.1 GI:5017510

KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 222)

AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

PUBMED 8889548

COMMENT

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa, 4156 MEBRF, Iowa City, IA 52242, USA

375 Newton Road,

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to verify it as a clone from the  
normalized ventricle at 13 dpc library cDNA Library Preparation:  
M.B. Soares Lab clone distribution: clones will be available  
through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=yes.

## FEATURES

source

1. 222  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-AGI-aao-b-11-0-UI"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-AGI"  
note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-AGI  
library is a normalized library constructed from 13 dpc  
rat ventricle. The tag is a string of 6 nucleotides  
present between the Not I site and the oligo-dT track.  
The library was constructed as described by Bonaldo,  
Lennon and Soares, Genome Research 6: 791-806, 1996.  
Tissue provided by Jim Lin, Department of Biology,  
University of Iowa.  
TAG\_TISSUE=ventricle at 13 dpc  
TAG\_LIB=UI-R-AGI  
TAG\_SEQ=CAGCGA"

## ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 222;  
Best Local Similarity 94.7%; Pred. No. 7.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGGCCCCAGCCCAAA 20

DB 73 GGGGGCCCCAGCCCAAA 55

## RESULT 6

AA657083/c

LOCUS

DEFINITION

AA657083 245 bp mRNA linear EST 04-NOV-1997  
vs23c10.r1 Barstead mouse myotubes MFLRBS Mus musculus cDNA clone  
IMAGS:1139058 5', mRNA sequence.

ACCESSION AA657083

VERSION AA657083.1 GI:2593237

KEYWORDS Est.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 245)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

REFERENCE The WashU-HMI Mouse EST Project

AUTHORS Unpublished (1996)

TITLE

JOURNAL

COMMENT

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnlnl.gov) for further information.  
MGI:620330

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 224.

## FEATURES

source

1. 245  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C3H"  
/db\_xref="taxon:10090"



PTT3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo.

ORIGIN

Query Match 87.0%; Score 17.4; DB 14; Length 283;  
Best Local Similarity 94.7%; Pred. No. 7.7e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCCACA 19  
|||||  
Db 167 CGAGGCGCCAGCCCCACA 149

RESULT 9  
AA623516/c  
LOCUS  
DEFINITION  
AA623516 304 bp mRNA linear EST 14-OCT-1997  
v33h05.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA  
clone IMAGE:1023033 5', mRNA sequence.

ACCESSION  
AA623516  
VERSION  
AA623516.1 GI:2527392  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)

REFERENCE  
AUTHORS  
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 304)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

TITLE  
JOURNAL  
COMMENT  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MG1:573809  
High quality sequence stop: 269.

FEATURES  
Location/Qualifiers  
1..304  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J x DBA/2J F1"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1023033"  
/tissue\_type="blastocyst"  
/dev\_stage="embryo (pre-implantation)"  
/lab\_hosts="DH10B"  
/clone\_lib="Knowles Solter mouse blastocyst B1"  
/note="Organ: embryo; Vector: pSPORT; Site: 1: Notr;  
Site 2: Sali; Cloned unidirectionally from mRNA prepared  
from 800 blastocysts. Primer: Sali (drt):  
5'-CGGTCGACCGCGACCGTTTITTTTTTTT-3'. cDNAs were  
cloned into the NotI/Sali sites of a pSPORT vector (Life  
Technologies). Two different size selections: B1 (larger  
inserts) and B3."

ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 304;  
Best Local Similarity 94.7%; Pred. No. 7.7e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCCACA 19  
|||||  
Db 91 CGAGGCGCCAGCCCCACA 73

RESULT 10  
AA144304/c  
LOCUS  
DEFINITION  
AA144304 305 bp mRNA linear EST 11-FEB-1997  
mr97f10.r1 Stratagene mouse embryonic carcinoma (#937317) Mus  
musculus cDNA clone IMAGE:605419 5', mRNA sequence.

ACCESSION  
AA144304  
VERSION  
AA144304.1 GI:1713674  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)

REFERENCE  
AUTHORS  
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 305)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

TITLE  
JOURNAL  
COMMENT  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MG1:370851  
Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 262.

FEATURES  
Location/Qualifiers  
1..305  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:605419"  
/tissue\_type="carcinoma"  
/dev\_stage="embryonic"  
/lab\_host="SOLR (kanamycin resistant)"  
/clone\_lib="Stratagene mouse embryonic carcinoma  
(#937317)"

/note="Vector: pBluescript SK-; Site: 1: EcoRI; Site: 2:  
XhoI; Cloned unidirectionally. Primer: Oligo dt. P19 cell  
line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor  
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3"

ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 305;  
Best Local Similarity 94.7%; Pred. No. 7.7e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCCACA 19  
|||||  
Db 181 CGAGGCGCCAGCCCCACA 163

RESULT 11  
BQ301646/c  
LOCUS  
DEFINITION  
BQ301646 308 bp mRNA linear EST 16-MAY-2002  
MRI-HB0017-151200-003-d07 HB0017 Homo sapiens cDNA, mRNA sequence.

ACCESSION  
BQ301646  
VERSION  
BQ301646.1 GI:20817168  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)

REFERENCE  
AUTHORS  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 308)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

10737800  
PUBMED

## COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?l=MR1&t2=MR1-HB0017-151200-003-d07&t3=2000-12-15&t4=1>)

Seq primer: puc 18 forward

High quality sequence stop: 6.

Location/Qualifiers

1..308

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/dev\_stage="Adult"

/clone\_lib="HB0017"

/note="Organ: bocio tumor; Vector: puc18; Site: 1: SmaI;

Site 2: SmaI; A mini-library was made by cloning products

derived from ORSTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

## FEATURES

source

ORIGIN  
Query Match 87.0%; Score 17.4; DB 13; Length 312;  
Best Local Similarity 94.7%; Pred. No. 7.7e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCCGAGCCCAACAA 20

|||||

Db 212 GAGGCGCCCGAGCCCAACAA 194

## ORIGIN

Query Match 87.0%; Score 17.4; DB 13; Length 308;

Best Local Similarity 94.7%; Pred. No. 7.7e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCCGAGCCCAACAA 20

|||||

Db 208 GAGGCGCCCGAGCCCAACAA 190

## RESULT 12

BQ301647/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

COMMENT

BQ301647 312 bp mRNA linear EST 16-MAY-2002

MR1-HB0017-151200-003-e03 HB0017 Homo sapiens cDNA, mRNA sequence.

BQ301647

BQ301647.1 GI:20817169

EST

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

COMMENT

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

10737800

PUBMED

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?l=MR1&t2=MR1-HB0017-151200-003-e03&t3=2000-12-15&t4=1>)

Seq primer: puc 18 forward

High quality sequence stop: 38.

Location/Qualifiers

1..312

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/dev\_stage="Adult"

/clone\_lib="HB0017"

/note="Organ: bocio tumor; Vector: puc18; Site: 1: SmaI;

Site 2: SmaI; A mini-library was made by cloning products

derived from ORSTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 13; Length 312;  
Best Local Similarity 94.7%; Pred. No. 7.7e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCCGAGCCCAACAA 20

|||||

Db 212 GAGGCGCCCGAGCCCAACAA 194

## RESULT 13

AI137544/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

10737800

PUBMED

Contact: Simpson A.J.G.

AI137544 316 bp mRNA linear EST 05-JUL-1999

UI-R-C2p-oc-g-03-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone

UI-R-C2p-oc-g-03-0-UI 3', mRNA sequence.

AI137544

AI137544.1 GI:3638321

EST

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 316)

Ronaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

The sequence tag present in the cDNA between the NotI site and the

oligo-dT track served to identify it as a clone from the normalized

adult Ovary library. cDNA Library Preparation: M. Faima Bonaldo,

Ph.D. Clone distribution: clones will be available through Research

Consortium at LNL ([info@image.lnl.gov](mailto:info@image.lnl.gov)). IMAGE ID=1787271

Seq primer: M13 Forward

POLYA=No.

```

FEATURES
  source
    Location/Qualifiers
      1..316
        /organism="Rattus norvegicus"
        /mol_type="mRNA"
        /strain="Sprague-Dawley"
        /db_xref="taxon:10116"
        /clone="UI-R-C2p-cc-g-03-0-U1"
        /dev_stage="adult"
        /lab_host="DH10B (Life Technologies)"
        /clone_lib="UI-R-C2p"
        /note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-C2p
library is a subtracted library derived from the UI-R-C1
library, which is a subtracted library derived from the
UI-R-C0 library. The UI-R-C0 library consisted of a
mixture of individually tagged normalized libraries
constructed from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
embryo. The tag is a string of 3-5 nucleotides present
between the Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture. The subtracted library (UI-R-C2p) was
constructed as follows: PCR amplified cDNA inserts from
UI-R-C1 clones from which 3' ESTs had been derived was
used as a driver in a hybridization with the UI-R-C1
library in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library) was
purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
UI-R-C2p library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research 6:
791-806, 1996)."

ORIGIN
  Query Match      87.0%; Score 17.4; DB 9; Length 316;
  Best Local Similarity 94.7%; Pred. No. 7.8e+03;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCCGCCGCCACAA 20
    |||||
DB 81 GGGGCGCCCGCCGCCACAA 63

RESULT 14
W98779/c
LOCUS      W98779
DEFINITION mg11h02.r1 Soares mouse embryo NM013.5 14.5 Mus musculus cDNA
ACCESSION W98779
VERSION   W98779.1 GI:1428687
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE 1 (bases 1 to 351)
AUTHORS   Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,
Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,
Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,
Theising B., Wylie T., Lennon G., Soares B., Wilson R. and
Waterston R.
TITLE     The WashU-HMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   Contact: Marra M/Mouse EST Project
          Washington University School of MedicineP
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: mouseest@watson.wustl.edu
          This clone is available royalty-free through LNL; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES
  source
    Location/Qualifiers
      1..363
        /organism="Mus musculus"
        /mol_type="mRNA"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="IMAGE:423507"
        /sex="unknown"
        /tissue_type="embryo"
        /dev_stage="13.5-14.5dpc total fetus"
        /lab_host="DH10B"
        /clone_lib="Soares mouse embryo NM013.5 14.5"
        /note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTTCAATCAATCGATGAGGCGCGCCGCCGCCGGAATTTTCTTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

ORIGIN
  Query Match      87.0%; Score 17.4; DB 14; Length 351;
  Best Local Similarity 94.7%; Pred. No. 7.9e+03;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCGCCACCA 19
    |||||
DB 206 CGAGGCGCCCGCCGCCACCA 188

RESULT 15
AA792229/c
LOCUS      AA792229
DEFINITION ve40b10.r1 Soares mammary_gland NDMMG Mus musculus cDNA clone
ACCESSION AA792229
VERSION   AA792229.1 GI:2855184
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE 1 (bases 1 to 363)
AUTHORS   Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,
Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,
Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,
Theising B., Wylie T., Lennon G., Soares B., Wilson R. and
Waterston R.
TITLE     The WashU-HMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   Contact: Marra M/Mouse EST Project
          Washington University School of MedicineP
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: mouseest@watson.wustl.edu
          This clone is available royalty-free through LNL; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES
  source
    Location/Qualifiers
      1..363
        /organism="Mus musculus"

```

```

/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:820603"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares mammary gland NbMMG"
/note="Organ: mammary gland; Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo (RT) primer 15.
TGTTACCACTCAGTAGTGCGGCGCCGCAAGTGTTTTTTTTTTTTTTTTTTT
T 3'; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fátima
Bonaldo."

```

## ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 363;  
Best Local Similarity 94.7%; Pred. No. 8e+03;  
Matches 18: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCCACA 19  
D'b 215 CGAGGCACCCAGCCCCACA 197

RESULT 16	W29177/c	LOCUS	W29177	364 bp	mRNA	linear	EST 11-SEP-1986
DEFINITION			mc2dc01.r1 Soares mouse p3NNF19.5		Mus musculus	CDNA clone	
IMAGE:			349248 5', mRNA sequence.				

ESL.  
SOURCE  
ORGANISM  
REFERENCE

Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 364)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisels,S., Kucaba,T., Lucy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steproe,M., Tan,F., Underwood,K., Moore,B.,  
Thaising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

WashU-HHMI Mouse ESR Project  
Washington University School  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)  
This clone is available royalty-free through LBNL ; contact the  
IMAGE Consortium ([info@image.lbl.gov](mailto:info@image.lbl.gov)) for further information.

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was amplified with a Not I - oligo(dT) primer [5', TGTACCACTCAAGTGGAGCGCGCGCATTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pWT3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 14; Length 364;  
Best Local Similarity 94.7%; Pred. No. 8e+03;  
Matches 18: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCCACA 19  
Db 255 CGAGGCACCCAGCCCCACA 237

<b>RESULT 17.</b>	
AA815612/c	
LOCUS	
DEFINITION	
AA815612	mRNA
vt03al.r1	370 bp
Bartstead mouse myotubes MFLRBS	linear
Mus musculus cDNA clone	EST 13-FEB-1998
IMAGE:1161980.5'	mRNA sequence.

KEYWORDS	ORGANISM	REFERENCE	AUTHORS
ESI.			
Mus musculus (house mouse)	Mus musculus		
	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
	1 (bases 1 to 370)		
	Marra, W., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,		
	Geisel, S., Kucaba, T., Lucy, M., Le, M., Martin, J., Morris, M.,		
	Schellenberg, K., Steptos, M., Tan, F., Underwood, K., Moore, B.,		
	Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and		
	Waterston, R.		

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouse@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
MGI:627892  
Seq primer: -28ml3 rev2 ET from Amersham  
Bio: [email:seq.sentence.prim.347](mailto:email:seq.sentence.prim.347)

```

FEATURES
source
High quality sequence scdp: 347.
location/Qualifiers
1..370
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C3H"
/db_xref="taxon:10090"
/clone="IMAGE:1161980"

```

/note="vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site1: EcoRI; Site 2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACGAATGAGTCAGCGCCGCCCTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATTCGATCTCTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead. The C3C12 cell line (available from ATCC, catalog # CRL-1772) differentiates rapidly, forming contractile myotubes and producing



	characteristic muscle proteins.	"	
ORIGIN			
Query Match	87.0%; Score 17.4; DB 9;	Length 370;	
Best Local Similarity	94.7%; Pred. No. 8e+03;		
Matches	18; Conservative 0; Mismatches 1;	Indels 0; Gaps 0;	
QY	1 CGAGGCGCCAGCCCCACA 19 		
Db	109 CGAGGCACCCAGCCCCACA 91 		
RESULT 18			
A1180711/c			
LOCUS	396 bp mRNA linear EST 08-OCT-1998		
DEFINITION	ub51b08.r1 Soares mammary gland MbMMG Mus musculus cDNA clone		
IMAGE:	1395831 5', mRNA sequence.		
ACCESSION	A1180711		
VERSION	A1180711.1 GI:3731349		
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 396)		
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,W., Tan,F., Underwood,K., Moore,B., Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
TITLE	The WashU-HHMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MG1:907547		
Seq primer:	-28ml3 rev2 ET from Amersham		
High quality sequence stop:	307.		
FEATURES			
source	Location/Qualifiers		
1..396	/organism="Mus musculus" /mol_type="mRNA" /strain="CS7BL/6J" /db_xref="taxon:10090" /clone="IMAGE:1395831" /sex="male" /tissue_type="mammary gland" /dev_stage="4 weeks" /lab_host="DH10B" /clone_lib="Soares mammary_gland_MbMMG" /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGAAGTCGGAGCGGCCGCAATGTGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Patima Bonaldo."		
Query Match	87.0%; Score 17.4; DB 9;	Length 396;	
Best Local Similarity	94.7%; Pred. No. 8.1e+03;		
Matches	18; Conservative 0; Mismatches 1;	Indels 0; Gaps 0;	
ORIGIN			
QY	1 CGAGGCGCCAGCCCCACA 19 		
Db	109 CGAGGCACCCAGCCCCACA 91 		
RESULT 18			
A1180711/c			
LOCUS	396 bp mRNA linear EST 08-OCT-1998		
DEFINITION	ub51b08.r1 Soares mammary_gland MbMMG Mus musculus cDNA clone		
IMAGE:	1395831 5', mRNA sequence.		
ACCESSION	A1180711		
VERSION	A1180711.1 GI:3731349		
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 396)		
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,W., Tan,F., Underwood,K., Moore,B., Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
TITLE	The WashU-HHMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MG1:907547		
Seq primer:	-28ml3 rev2 ET from Amersham		
High quality sequence stop:	307.		
FEATURES			
source	Location/Qualifiers		
1..396	/organism="Mus musculus" /mol_type="mRNA" /strain="CS7BL/6J" /db_xref="taxon:10090" /clone="IMAGE:1395831" /sex="male" /tissue_type="mammary gland" /dev_stage="4 weeks" /lab_host="DH10B" /clone_lib="Soares mammary_gland_MbMMG" /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGAAGTCGGAGCGGCCGCAATGTGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Patima Bonaldo."		
Query Match	87.0%; Score 17.4; DB 9;	Length 396;	
Best Local Similarity	94.7%; Pred. No. 8.1e+03;		
Matches	18; Conservative 0; Mismatches 1;	Indels 0; Gaps 0;	
ORIGIN			
QY	1 CGAGGCGCCAGCCCCACA 19 		
Db	109 CGAGGCACCCAGCCCCACA 91 		
RESULT 18			
A1180711/c			
LOCUS	396 bp mRNA linear EST 08-OCT-1998		
DEFINITION	ub51b08.r1 Soares mammary_gland MbMMG Mus musculus cDNA clone		
IMAGE:	1395831 5', mRNA sequence.		
ACCESSION	A1180711		
VERSION	A1180711.1 GI:3731349		
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 396)		
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,W., Tan,F., Underwood,K., Moore,B., Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
TITLE	The WashU-HHMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MG1:907547		
Seq primer:	-28ml3 rev2 ET from Amersham		
High quality sequence stop:	307.		
FEATURES			
source	Location/Qualifiers		
1..396	/organism="Mus musculus" /mol_type="mRNA" /strain="CS7BL/6J" /db_xref="taxon:10090"		

KEYWORDS  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM

REFERENCE  
AUTHORS Rattus norvegicus  
TITLE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

JOURNAL 1 (bases 1 to 434)  
MEDLINE Bonaldo,M.F., Lennon,G. and Soares,M.B.  
PUBMED Normalization and subtraction: two approaches to facilitate gene  
COMMENT discovery  
97044477  
8889548  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
The sequence contained an oligo-dt track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dt track served to verify it as a clone from the  
normalized rat eye library cDNA library Preparation: M.B. Soares  
Lab Clone distribution: clones will be available through Research  
Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLVA=Yes.

FEATURES  
source  
Location/Qualifiers  
1..434  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-CV1-brx-g-10-0-UI"  
/dev\_stage="ADULT"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-CV1"  
/note="Vector: pTT3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CV1  
library is a normalized library constructed from rat eye  
tissue. For a detailed description of the library from  
which this clone was derived, please visit our web site at  
rategen.uiowa.edu. The subtraction has been previously  
described in (Bonaldo, Lennon and Soares, Genome Research  
6:791-806, 1996)  
TAG\_TISSUE=rat eye  
TAG\_LIB=UI-R-CV1  
TAG\_SEQ=CAGCC"

ORIGIN  
Query Match 87.0%; Score 17.4; DB 12; Length 434;  
Best Local Similarity 94.7%; Pred. No. 8.3e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GAGGGCCCCCAGCCCCACAA 20  
|||||  
DB 82 GGGGGCCCCCAGCCCCACAA 64  
|||||

RESULT 21  
AA471939/c  
LOCUS  
DEFINITION AA471939 461 bp mRNA linear EST 18-JUN-1997  
v995h06.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA  
clone IMAGE:873755 5', mRNA sequence.  
ACCESSION AA471939  
VERSION AA471939.1 GI:2199930  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 461)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HHMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:513235  
Seq primer: -28ml3 rev2 ET from Amersham.

FEATURES  
source  
Location/Qualifiers  
1..461  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="TMAGE:873755"  
/sex="mixed"  
/tissue\_type="pooled organs"  
/dev\_stage="7 day"  
/lab\_host="DH10B"  
/clone\_lib="Barstead mouse pooled organs MPLRB4"  
/note="Organ: pooled; Vector: pTT3D-Pac (Pharmacia) with  
a modified polylinker; Site 1: EcoRI; Site 2: NotI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTAGCAATCTGAGAGGAGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
[GTGATTCGGTACC], digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pTT3 vector.  
Library constructed by Bob Barstead."

ORIGIN  
Query Match 87.0%; Score 17.4; DB 9; Length 461;  
Best Local Similarity 94.7%; Pred. No. 8.4e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CGAGGGCCCCCAGCCCCACA 19  
|||||  
DB 227 CGAGGGCCCCCAGCCCCACA 209  
|||||

RESULT 22  
W53227/c  
LOCUS  
DEFINITION W53227 507 bp mRNA linear EST 03-JUN-1996  
nd19a02.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA  
clone IMAGE:36858 5', mRNA sequence.  
ACCESSION W53227  
VERSION W53227.1 GI:1357053  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 507)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HHMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project

Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:230290

Seq primer: mob.REGA+ET

High quality sequence stop: 358.

#### FEATURES

Location/Qualifiers  
1..507

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:368858"

/sex="unknown"

/tissue\_type="embryo"

/dev\_stage="13.5-14.5dpc total fetus"

/lab\_host="DH10B"

/clone\_lib="Soares mouse embryo NbME13.5 14.5"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5,

TGTTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTT

T 3'], on equal amounts of mRNA from 2 13.5dpc and 2

14.5dpc embryos [total RNA provided by Minoru Ko, Wayne

State Univ., from 2 ]; double-stranded cDNA was ligated to

Eco RI adaptors (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of the modified

pT73 vector. Library went through one round of

normalization, and was constructed by Bento Soares and

M.Fatima Bonaldo. "

ORIGIN

Query Match 87.0%; Score 17.4; DB 14; Length 507;

Best Local Similarity 94.7%; Pred. No. 8.6e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGCGCCGCGCCGCCACA 19

Db 202 CGAGCACCAGCCGCCACA 184

RESULT 23

AA272464/c

LOCUS

DEFINITION

IMAGE:763438 5', mRNA sequence.

AA272464

VERSION

KEYWORDS

EST.

SOURCE

ORGANISM

Mus musculus (house mouse)

REFERENCE

AUTHORS

Waterston,R.

The WashU-HMM Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMM Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

ORIGIN

Query Match 87.0%; Score 17.4; DB 14; Length 507;

Best Local Similarity 94.7%; Pred. No. 8.6e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGCGCCGCGCCGCCACA 19

Db 202 CGAGCACCAGCCGCCACA 184

RESULT 23

AA272464/c

LOCUS

DEFINITION

IMAGE:763438 5', mRNA sequence.

AA272464

VERSION

KEYWORDS

EST.

SOURCE

ORGANISM

Mus musculus (house mouse)

REFERENCE

AUTHORS

Waterston,R.

The WashU-HMM Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMM Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:464358

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 491.

#### FEATURES

source

Location/Qualifiers

1..534

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:763438"

/sex="unknown"

/tissue\_type="fetus"

/dev\_stage="12.5dpc total fetus"

/lab\_host="DH10B"

/clone\_lib="Soares mouse 3NME12 5"

/note="Organ: whole fetus; Vector: pT73D-Pac (Pharmacia)

with a modified polylinker; Site 1: Not I; Site 2: Eco RI;

1st strand cDNA was primed with a Not I - oligo(dT) primer

[5, TGTTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTT

3'], on total mouse RNA [provided by Minoru Ko, Wayne

State Univ.]; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pT73 vector.

Library went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo. "

#### ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 534;

Best Local Similarity 94.7%; Pred. No. 8.7e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGCGCCGCGCCGCCACA 19

Db 190 CGAGCACCAGCCGCCACA 172

RESULT 24

BE395928/c

LOCUS

DEFINITION

562 bp mRNA linear EST 21-JUL-2000

BE395928

VERSION

KEYWORDS

EST.

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 562)

NIH-MGC http://imgc.ncbi.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: ARCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LHC351 row: 1 column: 02

High quality sequence stop: 562.

Location/Qualifiers

1..562

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:3659977"

/tissue\_type="endometrium, adenocarcinoma cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_44"

#### FEATURES

source

Location/Qualifiers

1..562

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:3659977"

/tissue\_type="endometrium, adenocarcinoma cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_44"

/note="Organ: uterus; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 562;  
Best Local Similarity 94.7%; Pred. No. 8.8e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCCGCCGACCA 20

Db 74 GAGGCGCCCGCCGACCA 56

## RESULT 25

AA458319/c

## LOCUS

DEFINITION VG48e07.r1 Soares mammary gland NbMMG Mus musculus cDNA clone

## ACCESSION

AA458319

## VERSION

AA458319.1

## KEYWORDS

EST.

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 588)

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:508700

Seq primer: -28mi3 rev2 ET from Amersham

High quality sequence stop: 450.

Location/Qualifiers

1..588

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:864612"

/sex="male"

/tissue\_type="mammary gland"

/development="4 weeks"

/lab\_host="DH10B"

/clone\_lib="Soares mammary gland NbMMG"

/note="Organ: mammary gland; Vector: pTV73D-Pac

(Pharmacia) with a modified polylinker; Site 1: Not I;

Site 2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo(dT) primer [5,

TGTTACCAATCTGAGTGGAGCGCCGCGATGTTTTTTTTTTTTTTTTTTT

T 3']; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pTV73 vector.

RNA provided by Dr. Minoru Ko, Wayne State Univ. Library

constructed and normalized by Bento Soares and M. Fatima

Bonaldo."

## FEATURES

source

## ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 588;  
Best Local Similarity 94.7%; Pred. No. 8.8e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCGACCA 19

Db 143 CGAGGCGCCCGCCGACCA 125

## RESULT 26

AA519009/c

## LOCUS

DEFINITION v109d10.r1 Barstead mouse myotubes MPLRBS Mus musculus cDNA clone

IMAGE:903283 5', mRNA sequence.

ACCESSION AA519009

VERSION AA519009.1

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 590)

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:5233947

Seq primer: -28mi3 rev2 ET from Amersham

High quality sequence stop: 489.

Location/Qualifiers

1..590

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C3H"

/db\_xref="taxon:10090"

/clone="IMAGE:903283"

/cell\_line="C2C12"

/lab\_host="DH10B"

/clone\_lib="Barstead mouse myotubes MPLRBS"

/note="Vector: pTV73D-Pac (Pharmacia) with a modified

polylinker; Site 1: EcoRI; Site 2: NotI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5,

TGTTACCAATCTGAGTGGAGCGCCGCGATGTTTTTTTTTTTTTTTTTTT

3']; double-stranded cDNA was ligated to Eco RI adaptors

[AATTCGATCCTTG], digested with Not I and cloned into the

Not I and Eco RI sites of the modified pTV73 vector.

Library constructed by Bob Barstead. The C2C12 cell line

(available from ATCC, catalog # CRL-1772) differentiates

rapidly, forming contractile myotubes and producing

characteristic muscle proteins."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 590;  
Best Local Similarity 94.7%; Pred. No. 8.9e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCGACCA 19

Db 208 CGAGGCGCCCGCCGACCA 190

RESULT 27  
 BI838149  
 LOCUS  
 DEFINITION 603083654P1 NIH\_MGC\_120 Homo sapiens cDNA clone IMAGE:5222917 5',  
 mRNA sequence.  
 ACCESSION BI838149  
 VERSION BI838149.1 GI:15949699  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 781)  
 NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM11560 row: k column: 14  
 High quality sequence stop: 342.  
 Location/Qualifiers  
 1..781  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5222917"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC 120"  
 /note="Organ: pCooled pancreas and spleen; Vector:  
 pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA  
 source anonymous pool of spleen and pancreas from 28 yo  
 male. Library is oligo-dT primed and directionally cloned  
 (EcoRV site is destroyed upon cloning). Average insert  
 size 1.5 kb, insert size range 1-2.5 kb. Library is  
 normalized and enriched for full-length clones and was  
 constructed by C. Gruber (Invitrogen). Research Genetics  
 tracking code 025. Note: this is a NIH\_MGC Library."

# FEATURES source

1..893  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultivar="Morex"  
 /sub\_species="vulgare"  
 /db\_xref="taxon:112509"  
 /clone="HVSMEF0021D17f"  
 /tissue\_type="Seedling root"  
 /lab\_host="HJG121"  
 /clone\_lib="Hordeum vulgare seedling root EST library  
 HVCNDA0007 (Etiolated and unstressed)"  
 /note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;  
 Seeds were surface sterilized then germinated under axenic  
 conditions in the dark at room temperature on filter paper  
 with water, nystatin and cefotaxime in covered  
 crystallization dishes. Five-day old seedling roots were  
 then harvested, total RNA was prepared, poly(A) RNA was  
 purified, one primary unamplified cDNA library was made,  
 and 1 million pfu were in vivo excised to give pBluescript  
 SK(-) cDNA phagemids. These steps were performed in the TJ  
 Close laboratory at the University of California,  
 Riverside (Choi, Close, Fenton). Phagemids were plated and  
 picked at the Clemson University Genomics Institute (CUGI)  
 (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA  
 preparations, DNA sequencing and sequence analysis were  
 performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,  
 Oates, Rambo, Main). The sequence has been trimmed to  
 remove vector sequence and contains a minimum of 100 bases  
 of pred value 20 or above. For more details on library  
 preparation and sequence analysis see  
 http://www.genome.clemson.edu/projects/barley. To order  
 this clone see http://www.genome.clemson.edu/orders/Also  
 see Close TJ, Wing R, Kleinhofs A, Wise R (2001)  
 Genetically and physically anchored EST resources for  
 barley genomics. Barley Genetics Newsletter 31:29-30.  
 (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

# ORIGIN

Query Match 87.0%; Score 17.4; DB 12; Length 781;  
 Best Local Similarity 94.7%; Pred. No. 9.4e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2 GAGCGGCCAGCCACAA 20  
 |||||  
 Db 695 GAGCGGCTCAGCCACAA 713  
 RESULT 28  
 BI260146  
 LOCUS  
 DEFINITION HVSMF0021D17f Hordeum vulgare seedling root EST library HVCNDA0007  
 (Etiolated and unstressed) Hordeum vulgare subsp. vulgare cDNA  
 clone HVSMEF0021D17f, mRNA sequence.  
 ACCESSION BI260146  
 VERSION BI260146.3 GI:16316189  
 KEYWORDS EST.  
 SOURCE Hordeum vulgare subsp. vulgare  
 ORGANISM Hordeum vulgare subsp. vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooideae; Triticeae; Hordeum.  
 1 (bases 1 to 893)  
 Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D.,

# TITLE JOURNAL COMMENT

Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W.,  
 Fenton, R.D., Oates, R. and Main, D.  
 Development of a genetically and physically anchored EST resource  
 for barley genomics: Morex unstressed seedling root cDNA library  
 Unpublished (2001)  
 On Nov 16, 2000 this sequence version replaced gi:13120492.  
 Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: twing@clemson.edu  
 Total hg bases = 366  
 Seq primer: AATAACCTCACTAAAGG  
 High quality sequence stop: 753.  
 Location/Qualifiers  
 1..893  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultivar="Morex"  
 /sub\_species="vulgare"  
 /db\_xref="taxon:112509"  
 /clone="HVSMEF0021D17f"  
 /tissue\_type="Seedling root"  
 /lab\_host="HJG121"  
 /clone\_lib="Hordeum vulgare seedling root EST library  
 HVCNDA0007 (Etiolated and unstressed)"  
 /note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;  
 Seeds were surface sterilized then germinated under axenic  
 conditions in the dark at room temperature on filter paper  
 with water, nystatin and cefotaxime in covered  
 crystallization dishes. Five-day old seedling roots were  
 then harvested, total RNA was prepared, poly(A) RNA was  
 purified, one primary unamplified cDNA library was made,  
 and 1 million pfu were in vivo excised to give pBluescript  
 SK(-) cDNA phagemids. These steps were performed in the TJ  
 Close laboratory at the University of California,  
 Riverside (Choi, Close, Fenton). Phagemids were plated and  
 picked at the Clemson University Genomics Institute (CUGI)  
 (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA  
 preparations, DNA sequencing and sequence analysis were  
 performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,  
 Oates, Rambo, Main). The sequence has been trimmed to  
 remove vector sequence and contains a minimum of 100 bases  
 of pred value 20 or above. For more details on library  
 preparation and sequence analysis see  
 http://www.genome.clemson.edu/projects/barley. To order  
 this clone see http://www.genome.clemson.edu/orders/Also  
 see Close TJ, Wing R, Kleinhofs A, Wise R (2001)  
 Genetically and physically anchored EST resources for  
 barley genomics. Barley Genetics Newsletter 31:29-30.  
 (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

# FEATURES source

1..893  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultivar="Morex"  
 /sub\_species="vulgare"  
 /db\_xref="taxon:112509"  
 /clone="HVSMEF0021D17f"  
 /tissue\_type="Seedling root"  
 /lab\_host="HJG121"  
 /clone\_lib="Hordeum vulgare seedling root EST library  
 HVCNDA0007 (Etiolated and unstressed)"  
 /note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;  
 Seeds were surface sterilized then germinated under axenic  
 conditions in the dark at room temperature on filter paper  
 with water, nystatin and cefotaxime in covered  
 crystallization dishes. Five-day old seedling roots were  
 then harvested, total RNA was prepared, poly(A) RNA was  
 purified, one primary unamplified cDNA library was made,  
 and 1 million pfu were in vivo excised to give pBluescript  
 SK(-) cDNA phagemids. These steps were performed in the TJ  
 Close laboratory at the University of California,  
 Riverside (Choi, Close, Fenton). Phagemids were plated and  
 picked at the Clemson University Genomics Institute (CUGI)  
 (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA  
 preparations, DNA sequencing and sequence analysis were  
 performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,  
 Oates, Rambo, Main). The sequence has been trimmed to  
 remove vector sequence and contains a minimum of 100 bases  
 of pred value 20 or above. For more details on library  
 preparation and sequence analysis see  
 http://www.genome.clemson.edu/projects/barley. To order  
 this clone see http://www.genome.clemson.edu/orders/Also  
 see Close TJ, Wing R, Kleinhofs A, Wise R (2001)  
 Genetically and physically anchored EST resources for  
 barley genomics. Barley Genetics Newsletter 31:29-30.  
 (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

# ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 893;  
 Best Local Similarity 94.7%; Pred. No. 9.7e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 CGAGCGGCCAGCCACACA 19  
 |||||  
 Db 691 CGAGCGGCCAGCCACACA 709  
 RESULT 29  
 BZ569703  
 LOCUS  
 DEFINITION pacs2-164\_979.s1 pacs2-164 Pseudomonas aeruginosa genomic clone  
 pacs2-164\_979, genomic survey sequence.  
 ACCESSION BZ569703  
 VERSION BZ569703.1 GI:27204692  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa

```

ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1185)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-genome sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
    source
        Location/Qualifiers
            1..1185
                /organism="Pseudomonas aeruginosa"
                /mol_type="genomic DNA"
                /strain="2-164"
                /db_xref="taxon:287"
                /clone="pacs2-164_979"
                /clone_lib="pacs2-164"
                /note="clinical isolate 2-164 Whole genomic shotgun
                library."

ORIGIN
    Query Match 87.0%; Score 17.4; DB 28; Length 1185;
    Best Local Similarity 94.7%; Pred. No. 1e+04;
    Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCGACCA 19
    |||||
Db 614 CGAGGCGCCCGCCGACCA 632

RESULT 30
AGI83774/c 1229 bp DNA linear GSS 09-JAN-2002
LOCUS Pan troglodytes DNA, clone: RP43-057F18.TJ, genomic survey
DEFINITION sequence.
ACCESSION AGI83774
VERSION AGI83774.1 GI:16713454
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE BAC end sequences of Library RPCI-43
JOURNAL Unpublished
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Sueni-ro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpansegsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.

```

```

FEATURES
    source
        Location/Qualifiers
            1..1229
                /organism="Pan troglodytes"
                /mol_type="genomic DNA"
                /db_xref="taxon:9598"
                /clone="RP43-057F18.TJ"
                /sex="male"
                /cell_type="lymphocytes"
                /clone_lib="RPCI-43 Chimpanzee Male BAC Library"

ORIGIN
    Query Match 87.0%; Score 17.4; DB 29; Length 1229;
    Best Local Similarity 94.7%; Pred. No. 1e+04;
    Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCGACCA 19
    |||||
Db 976 CGAGGCGCCCGCCGACCA 958

Search completed: June 20, 2004, 14:13:08
Job time : 982.603 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 22.1164 Seconds  
(without alignments)  
501.846 Million cell updates/sec

Title: US-10-624-714-11

Perfect score: 20

Sequence: 1 cgaggcgccagcccccacaa 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTCUTS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	95.0	4403765	3	US-09-103-840A-2
2	19	95.0	4411529	3	US-09-103-840A-1
3	16	80.0	3845	2	US-08-220-240A-1
4	15.8	79.0	477	4	US-09-621-976-18886
5	15.8	79.0	947	4	US-09-347-878-49
6	15.8	79.0	981	3	US-08-815-225-1
7	15.8	79.0	4380	4	US-08-955-565A-3
8	15.8	79.0	13857	4	US-09-620-312D-75
9	15.8	79.0	4403765	3	US-09-103-840A-2
10	15.8	79.0	4411529	3	US-09-103-840A-1
11	15.4	77.0	216	4	US-09-252-991A-1061
12	15.4	77.0	432	4	US-09-252-991A-3530
13	15.4	77.0	504	4	US-09-252-991A-3564
14	15.4	77.0	636	4	US-09-252-991A-7251
15	15.4	77.0	708	4	US-09-252-991A-13744
16	15.4	77.0	804	4	US-09-252-991A-13605
17	15.4	77.0	807	4	US-09-252-991A-13398
18	15.4	77.0	858	4	US-09-033-556-8
19	15.4	77.0	858	4	US-09-474-699-12
20	15.4	77.0	858	4	US-09-151-376-8
21	15.4	77.0	963	4	US-09-252-991A-7371
22	15.4	77.0	984	4	US-09-252-991A-13494
23	15.4	77.0	1002	4	US-09-252-991A-3539
24	15.4	77.0	1143	4	US-09-252-991A-949
25	15.4	77.0	1350	4	US-09-252-991A-1098
26	15.4	77.0	1536	4	US-09-489-039A-1997
27	15.4	77.0	1644	4	US-09-252-991A-3548

Sequence 1134, Ap  
Sequence 1, Appli  
Sequence 188, App  
Sequence 3523, Ap  
Sequence 1, Appli  
Sequence 21, Appli  
Sequence 3, Appli  
Sequence 649, App  
Sequence 246, App  
Sequence 892, App  
Sequence 1139, Ap  
Sequence 944, App  
Sequence 20, Appli  
Sequence 1, Appli  
Sequence 10, Appli  
Sequence 10, Appli  
Sequence 10, Appli  
Sequence 4030, Ap  
Sequence 1516, Ap  
Sequence 1557, Ap  
Sequence 1662, Ap  
Sequence 2928, Ap  
Sequence 1466, Ap  
Sequence 3025, Ap  
Sequence 305, App  
Sequence 6706, Ap  
Sequence 817, App  
Sequence 1619, Ap  
Sequence 973, App  
Sequence 3035, Ap  
Sequence 929, App  
Sequence 143, App  
Sequence 506, App  
Sequence 8721, Ap  
Sequence 1969, Ap  
Sequence 44, Appli  
Sequence 8940, Ap  
Sequence 6673, Ap  
Sequence 8823, Ap  
Sequence 1621, Ap  
Sequence 251, App  
Sequence 7, Appli  
Sequence 1, Appli  
Sequence 1620, Ap  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 281, App  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 9144, Ap  
Sequence 6641, Ap  
Sequence 1, Appli  
Sequence 909, App  
Sequence 202, App  
Sequence 9, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 12, Appli  
Sequence 13, Appli  
Sequence 7, Appli  
Sequence 5, Appli  
Sequence 3, Appli  
Sequence 32, Appli  
Sequence 3, Appli  
Sequence 3, Appli

101	14.4	72.0	28	2	US-08-859-998-304	Sequence 304, App	174	14.2	71.0	2358	1	US-08-339-152A-15	Sequence 15, Appl
102	14.4	72.0	28	4	US-09-225-928-304	Sequence 304, App	175	14.2	71.0	2358	2	US-08-007-999B-2	Sequence 2, Appl
103	14.4	72.0	28	4	US-09-225-201B-304	Sequence 304, App	176	14.2	71.0	2358	2	US-08-689-276A-2	Sequence 2, Appl
104	14.4	72.0	261	4	US-09-252-991A-3723	Sequence 3723, App	177	14.2	71.0	2710	4	US-09-023-655-928	Sequence 928, App
105	14.4	72.0	384	4	US-09-280-116-236	Sequence 236, App	178	14.2	71.0	2710	4	US-09-620-312D-358	Sequence 358, App
106	14.4	72.0	693	4	US-09-402-515A-17	Sequence 17, Appl	179	14.2	71.0	3494	3	US-09-139-802-200	Sequence 200, App
107	14.4	72.0	700	3	US-08-998-416-712	Sequence 712, App	180	14.2	71.0	3494	4	US-09-653-786-200	Sequence 200, App
108	14.4	72.0	906	4	US-08-998-416-712	Sequence 712, App	181	14.2	71.0	3494	4	US-09-023-655-1006	Sequence 1006, App
109	14.4	72.0	921	4	US-09-091-952A-10821	Sequence 10821, A	182	14.2	71.0	3568	3	US-09-103-510B-1	Sequence 1, Appl
110	14.4	72.0	1011	4	US-09-091-952A-10821	Sequence 7, Appl	183	14.2	71.0	3568	4	US-09-492-639A-1	Sequence 1, Appl
111	14.4	72.0	1161	1	US-08-383-750-5	Sequence 383, App	184	14.2	71.0	3568	4	US-09-491-971-1	Sequence 1, Appl
112	14.4	72.0	1161	2	US-08-684-687-1	Sequence 5, Appl	185	14.2	71.0	3568	4	US-09-491-970A-1	Sequence 1, Appl
113	14.4	72.0	1161	3	US-08-684-687-1	Sequence 1, Appl	186	14.2	71.0	3568	4	US-08-164-292B-1	Sequence 1, Appl
114	14.4	72.0	1161	4	US-08-352-678-5	Sequence 5, Appl	187	14.2	71.0	4060	1	US-08-164-292B-3	Sequence 3, Appl
115	14.4	72.0	1161	4	US-09-536-954-5	Sequence 5, Appl	188	14.2	71.0	4060	1	US-08-164-292B-5	Sequence 5, Appl
116	14.4	72.0	1161	4	US-09-016-434-1440	Sequence 1440, App	189	14.2	71.0	4060	1	US-08-164-292B-7	Sequence 7, Appl
117	14.4	72.0	1161	4	US-09-023-655-1460	Sequence 1460, App	190	14.2	71.0	4060	1	US-08-845-623-1	Sequence 1, Appl
118	14.4	72.0	1164	5	PCT-US93-09636-5	Sequence 5, Appl	191	14.2	71.0	4060	3	US-08-845-623-3	Sequence 3, Appl
119	14.4	72.0	1173	4	US-09-489-039A-2704	Sequence 2704, App	192	14.2	71.0	4060	3	US-08-845-623-5	Sequence 5, Appl
120	14.4	72.0	1236	4	US-09-489-039A-3754	Sequence 3754, App	193	14.2	71.0	4060	3	US-08-845-623-7	Sequence 7, Appl
121	14.4	72.0	1263	4	US-09-252-991A-10863	Sequence 10863, A	194	14.2	71.0	4060	3	US-08-815-927-1	Sequence 1, Appl
122	14.4	72.0	1290	4	US-09-252-991A-3792	Sequence 3792, App	195	14.2	71.0	4060	3	US-08-815-927-3	Sequence 3, Appl
123	14.4	72.0	1312	1	US-08-205-506A-1	Sequence 1, Appl	196	14.2	71.0	4060	3	US-08-815-927-5	Sequence 5, Appl
124	14.4	72.0	1312	5	PCT-US94-02389-1	Sequence 1, Appl	197	14.2	71.0	4060	4	US-08-815-927-7	Sequence 7, Appl
125	14.4	72.0	1375	4	US-09-372-432A-37	Sequence 37, Appl	198	14.2	71.0	4060	4	US-09-103-330-1	Sequence 1, Appl
126	14.4	72.0	1566	2	US-08-405-175A-10	Sequence 10, Appl	199	14.2	71.0	4060	4	US-09-103-330-3	Sequence 3, Appl
127	14.4	72.0	1797	4	US-09-832-441-1	Sequence 1, Appl	200	14.2	71.0	4060	4	US-09-103-330-5	Sequence 5, Appl
128	14.4	72.0	1797	4	US-09-832-441-1	Sequence 1, Appl	201	14.2	71.0	4060	4	US-09-435-242-1	Sequence 1, Appl
129	14.4	72.0	1956	4	US-09-715-858-1	Sequence 1, Appl	202	14.2	71.0	4060	4	US-09-435-242-3	Sequence 3, Appl
130	14.4	72.0	2191	3	US-09-039-555B-12	Sequence 12, Appl	203	14.2	71.0	4060	4	US-09-435-242-5	Sequence 5, Appl
131	14.4	72.0	2742	4	US-09-252-991A-13678	Sequence 3678, App	204	14.2	71.0	4060	4	US-09-435-242-7	Sequence 7, Appl
132	14.4	72.0	2748	4	US-09-252-991A-13820	Sequence 13820, App	205	14.2	71.0	4060	4	US-09-435-242-9	Sequence 9, Appl
133	14.4	72.0	2784	1	US-08-262-465-3	Sequence 3, Appl	206	14.2	71.0	5398	4	US-08-426-630-1	Sequence 1, Appl
134	14.4	72.0	3318	4	US-08-405-254-7	Sequence 7, Appl	207	14.2	71.0	8056	3	US-09-136-608-14	Sequence 14, Appl
135	14.4	72.0	3472	6	5244792-1	Patent No. 5244792	208	14.2	71.0	8078	3	US-08-870-128-12	Sequence 12, Appl
136	14.4	72.0	3720	4	US-09-252-991A-13627	Sequence 13627, A	209	14.2	71.0	8078	4	US-09-445-247-12	Sequence 12, Appl
137	14.4	72.0	3993	4	US-09-167-109-5	Sequence 5, Appl	210	14.2	71.0	8082	1	US-08-306-691B-41	Sequence 41, Appl
138	14.4	72.0	4338	1	US-08-015-986A-1	Sequence 1, Appl	211	14.2	71.0	8082	1	US-08-187-785-1	Sequence 1, Appl
139	14.4	72.0	4338	2	US-08-448-363-1	Sequence 1, Appl	212	14.2	71.0	8082	11	US-09-167-322-11	Sequence 11, Appl
140	14.4	72.0	4700	3	US-09-150-460B-9	Sequence 9, Appl	213	14.2	71.0	8082	5	PCT-US93-06251-28	Sequence 28, Appl
141	14.4	72.0	7304	4	US-09-453-702B-174	Sequence 174, App	214	14.2	71.0	11558	5	PCT-US93-06251-23	Sequence 23, Appl
142	14.4	72.0	8065	4	US-09-091-952A-6	Sequence 6, Appl	215	14.2	71.0	14985	1	US-08-652-972A-6	Sequence 6, Appl
143	14.4	72.0	51259	3	US-08-781-891-209	Sequence 209, App	216	14.2	71.0	14985	5	PCT-US96-06231A-6	Sequence 35, Appl
144	14.4	72.0	51259	4	US-09-618-166-209	Sequence 209, App	217	14.2	71.0	34446	4	US-09-103-330-35	Sequence 35, Appl
145	14.4	72.0	536165	4	US-09-214-808-1	Sequence 1, Appl	218	14.2	71.0	34446	4	US-09-103-330-35	Sequence 35, Appl
146	14.2	71.0	201	4	US-09-252-991A-14528	Sequence 14528, A	219	14.2	71.0	70000	4	US-09-851-896-3	Sequence 3, Appl
147	14.2	71.0	276	4	US-09-252-991A-7396	Sequence 7396, App	220	14.2	71.0	80161	3	US-09-036-987A-1	Sequence 1, Appl
148	14.2	71.0	435	4	US-09-252-991A-4662	Sequence 4662, App	221	14.2	71.0	80161	3	US-09-370-700-1	Sequence 1, Appl
149	14.2	71.0	465	4	US-08-426-630-11	Sequence 11, Appl	222	14.2	71.0	80161	4	US-09-603-207-1	Sequence 1, Appl
150	14.2	71.0	543	4	US-09-252-991A-4453	Sequence 4453, App	223	14.2	71.0	536165	4	US-09-214-808-1	Sequence 1, Appl
151	14.2	71.0	549	3	US-08-851-190-2	Sequence 2, Appl	224	14	70.0	450	4	US-09-621-976-1037	Sequence 1037, App
152	14.2	71.0	603	4	US-09-252-991A-4154	Sequence 4154, App	225	14	70.0	2112	3	US-08-845-546-11	Sequence 11, Appl
153	14.2	71.0	655	3	US-08-548-809A-25	Sequence 25, Appl	226	14	70.0	2437	4	US-09-833-381-1478	Sequence 1478, App
154	14.2	71.0	728	4	US-09-972-800A-68	Sequence 68, Appl	227	14	70.0	2437	4	US-09-833-381-1479	Sequence 1479, App
155	14.2	71.0	783	4	US-09-252-991A-14661	Sequence 14661, A	228	14	70.0	2900	4	US-09-620-312D-1045	Sequence 1045, App
156	14.2	71.0	928	4	US-09-216-393B-143	Sequence 143, App	229	14	70.0	3650	4	US-09-620-312D-30	Sequence 30, Appl
157	14.2	71.0	1134	4	US-09-313-283C-359	Sequence 359, App	230	14	70.0	6763	2	US-08-756-506-23	Sequence 23, Appl
158	14.2	71.0	1149	4	US-09-252-991A-14783	Sequence 14783, A	231	14	70.0	10807	2	US-08-206-176-7	Sequence 7, Appl
159	14.2	71.0	1440	4	US-09-023-655-437	Sequence 427, App	232	14	70.0	10807	2	US-08-756-506-5	Sequence 5, Appl
160	14.2	71.0	1515	4	US-09-489-039A-6745	Sequence 6745, App	233	14	70.0	152331	3	US-09-128-155-16	Sequence 16, Appl
161	14.2	71.0	1539	3	US-08-818-112-32	Sequence 32, Appl	234	14	70.0	176373	3	US-09-128-155-17	Sequence 17, Appl
162	14.2	71.0	1539	4	US-08-818-111-32	Sequence 32, Appl	235	13.8	69.0	105	3	US-08-717-294-87	Sequence 87, Appl
163	14.2	71.0	1539	4	US-09-056-556-32	Sequence 32, Appl	236	13.8	69.0	113	1	US-08-255-471-3	Sequence 3, Appl
164	14.2	71.0	1539	4	US-09-073-596-32	Sequence 32, Appl	237	13.8	69.0	113	1	US-08-255-471-6	Sequence 6, Appl
165	14.2	71.0	1539	4	US-09-072-967-32	Sequence 32, Appl	238	13.8	69.0	113	1	US-08-255-471-7	Sequence 7, Appl
166	14.2	71.0	1702	4	US-09-023-655-430	Sequence 420, App	239	13.8	69.0	270	4	US-09-489-039A-2576	Sequence 2576, App
167	14.2	71.0	1707	4	US-09-023-655-449	Sequence 449, App	240	13.8	69.0	285	4	US-09-016-434-172	Sequence 172, App
168	14.2	71.0	1711	4	US-09-492-639A-7	Sequence 7, Appl	241	13.8	69.0	285	4	US-09-621-976-7806	Sequence 7806, App
169	14.2	71.0	1711	4	US-09-491-971-7	Sequence 7, Appl	242	13.8	69.0	317	4	US-09-008-097-1	Sequence 1, Appl
170	14.2	71.0	1711	4	US-09-491-970A-7	Sequence 7, Appl	243	13.8	69.0	334	4	US-09-621-976-11707	Sequence 11707, A
171	14.2	71.0	1864	1	US-08-454-720A-38	Sequence 38, Appl	244	13.8	69.0	420	3	US-08-943-731-198	Sequence 198, App
172	14.2	71.0	1866	4	US-09-489-039A-6738	Sequence 6738, App	245	13.8	69.0	436	4	US-09-621-976-10893	Sequence 10893, A
173	14.2	71.0	2187	4	US-09-302-769-7	Sequence 7, Appl	246	13.8	69.0				



```
C 247 13.8 69.0 439 4 US-09-621-376-9244
C 248 13.8 69.0 452 4 US-09-621-376-9111
C 249 13.8 69.0 468 4 US-09-621-376-3265
C 250 13.8 69.0 488 4 US-09-280-116-109
C 251 13.8 69.0 522 4 US-09-621-376-8994
C 252 13.8 69.0 591 2 US-08-481-814A-4
C 253 13.8 69.0 593 4 US-09-517-431E-10
C 254 13.8 69.0 618 4 US-09-489-039A-2443
C 255 13.8 69.0 680 4 US-09-833-381-976
C 256 13.8 69.0 699 3 US-08-998-416-930
C 257 13.8 69.0 764 3 US-09-188-930-263
C 258 13.8 69.0 764 4 US-09-724-864-66
C 259 13.8 69.0 764 4 US-09-312-283C-263
C 260 13.8 69.0 766 3 US-09-188-930-38
C 261 13.8 69.0 766 4 US-09-312-283C-38
C 262 13.8 69.0 771 4 US-09-489-039A-4829
C 263 13.8 69.0 789 4 US-09-252-991A-12447
C 264 13.8 69.0 798 4 US-09-252-991A-12808
C 265 13.8 69.0 877 4 US-09-204-865-7
C 266 13.8 69.0 900 4 US-09-489-039A-5980
C 267 13.8 69.0 960 4 US-09-489-039A-308
C 268 13.8 69.0 971 4 US-09-023-655-646
C 269 13.8 69.0 972 4 US-09-252-991A-9101
C 270 13.8 69.0 1008 4 US-09-252-991A-10020
C 271 13.8 69.0 1029 4 US-09-252-991A-15460
C 272 13.8 69.0 1044 4 US-09-252-991A-2435
C 273 13.8 69.0 1062 4 US-09-244-111-1
C 274 13.8 69.0 1098 4 US-09-252-991A-15400
C 275 13.8 69.0 1100 2 US-08-481-814A-5
C 276 13.8 69.0 1107 4 US-09-173-300-50
C 277 13.8 69.0 1134 4 US-09-252-991A-7137
C 278 13.8 69.0 1170 4 US-09-252-991A-7929
C 279 13.8 69.0 1194 4 US-09-252-991A-11352
C 280 13.8 69.0 1248 4 US-09-252-991A-7361
C 281 13.8 69.0 1294 2 US-08-665-647-4
C 282 13.8 69.0 1308 4 US-09-489-039A-2811
C 283 13.8 69.0 1314 4 US-09-242-737-1
C 284 13.8 69.0 1363 1 US-08-776-088-21
C 285 13.8 69.0 1363 5 PCT-US95-09145A-21
C 286 13.8 69.0 1377 4 US-09-252-991A-2251
C 287 13.8 69.0 1404 4 US-09-252-991A-7623
C 288 13.8 69.0 1404 4 US-09-489-039A-2728
C 289 13.8 69.0 1434 4 US-09-489-039A-2766
C 290 13.8 69.0 1461 4 US-09-894-844-84
C 291 13.8 69.0 1488 4 US-09-252-991A-15378
C 292 13.8 69.0 1500 4 US-09-252-991A-11312
C 293 13.8 69.0 1509 4 US-09-489-039A-6014
C 294 13.8 69.0 1542 3 US-09-008-271A-13
C 295 13.8 69.0 1551 4 US-09-489-039A-4788
C 296 13.8 69.0 1663 4 US-09-312-283C-370
C 297 13.8 69.0 1663 4 US-09-312-283C-416
C 298 13.8 69.0 1829 4 US-09-620-312D-274
C 299 13.8 69.0 1830 4 US-09-252-991A-3164
C 300 13.8 69.0 1842 4 US-09-252-991A-15508
```

## ALIGNMENTS

```
RESULT 1
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
```

```
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 95.0%; Score 19; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GAGGCGCCCGAGCCGCCACAA 20
DB 2599945 GAGGCGCCCGAGCCGCCACAA 2599963

RESULT 2
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match 95.0%; Score 19; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 GAGGCGCCCGAGCCGCCACAA 20
DB 2604135 GAGGCGCCCGAGCCGCCACAA 2604153
```

```
RESULT 3
US-08-220-240A-4/c
; Sequence 4, Application US/08220240A
; Patent No. 5955291
; GENERAL INFORMATION:
; APPLICANT: Alitalo, Kari
; APPLICANT: Matikainen, Marja-Terttu
; APPLICANT: Partanen, Juhna
; APPLICANT: Mäkelä, Tomi
; APPLICANT: Korhonen, Jaana
; TITLE OF INVENTION: ANTIBODIES RECOGNIZING TIE RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASE AND USES THEREOF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
```



```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..801
; LENGTH: 19..801
US-08-815-225-1

Query Match          79.0%; Score 15.8; DB 3; Length 981;
Best Local Similarity 89.5%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGGCGCCCGCCGCCACAA 20
   ||||| ||||| ||||| |||||
DB 125 GAGGCTCCCTGCCGCCACAA 107

RESULT 7
US-08-955-565A-3/c
; Sequence 3, Application US/08955565A
; Patent No. 6331388
; GENERAL INFORMATION:
; APPLICANT: Malkovsky, Miroslav
; TITLE OF INVENTION: Immune Response Enhancer Therapy
; FILE REFERENCE: WARF-02625
; CURRENT APPLICATION NUMBER: US/08/955,565A
; CURRENT FILING DATE: 1997-10-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 4380
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-08-955-565A-3

Query Match          79.0%; Score 15.8; DB 4; Length 4380;
Best Local Similarity 89.5%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGGCGCCCGCCGCCACAA 20
   ||||| ||||| ||||| |||||
DB 1929 GAGGAGCCCGGCCGCCACAA 1911

RESULT 8
US-09-620-312D-75
; Sequence 75, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John, Tillinghast
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19

```

```

; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 75
; LENGTH: 13857
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(13857)
US-09-620-312D-75

Query Match          79.0%; Score 15.8; DB 4; Length 13857;
Best Local Similarity 89.5%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCGCCACAA 19
   ||||| ||||| ||||| |||||
DB 4956 CGAGACCCCGCCGCCACAA 4974

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          79.0%; Score 15.8; DB 3; Length 4403765;
Best Local Similarity 89.5%; Pred. No. 98;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGGCGCCCGCCGCCACAA 20
   ||||| ||||| ||||| |||||
DB 531725 GAGGAGCCCGGCCGCCACAA 531707

RESULT 10
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24

```

```
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match          79.0%; Score 15.8; DB 3; Length 4411529;
Best Local Similarity 89.5%; Pred. No. 98;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGGCGCCCGCCGCCCCA 20
DB 530283 GAGGAGCCCGCCGCCCCA 530265

RESULT 11
US-09-252-991A-1061/c
; Sequence 1061, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1061
; LENGTH: 216
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1061

Query Match          77.0%; Score 15.4; DB 4; Length 216;
Best Local Similarity 94.1%; Pred. No. 3.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCGCCCCA 17
DB 207 CGTGGCGCCCGCCGCCCCA 191

RESULT 12
US-09-252-991A-3530/c
; Sequence 3530, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3530
; LENGTH: 432
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3530

Query Match          77.0%; Score 15.4; DB 4; Length 432;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;

US-09-252-991A-7251/c
; Sequence 7251, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7251
; LENGTH: 636
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7251

Query Match          77.0%; Score 15.4; DB 4; Length 636;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCGCCCCA 17
DB 131 CGTGGCGCCCGCCGCCCCA 147

RESULT 14
US-09-252-991A-7251/c
; Sequence 7251, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7251
; LENGTH: 636
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7251

Query Match          77.0%; Score 15.4; DB 4; Length 636;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCGCCCCA 17
DB 168 CAAGCGCCCGCCGCCCCA 152

RESULT 15
US-09-252-991A-13744
; Sequence 13744, Application US/09252991A
; Patent No. 6551795
```

```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13744
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13744

Query Match 77.0%; Score 15.4; DB 4; Length 708;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCCA 17
DB 460 CGCGCGCCCGCCCA 476

RESULT 16
US-09-252-991A-13605/c
; Sequence 13605, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13605
; LENGTH: 804
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13605

Query Match 77.0%; Score 15.4; DB 4; Length 804;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCCA 17
DB 286 CGCGCGCCCGCCCA 270

RESULT 17
US-09-252-991A-13398/c
; Sequence 13398, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13398
; LENGTH: 807
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13398

Query Match 77.0%; Score 15.4; DB 4; Length 807;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCCA 17
DB 717 CGCGCGCCCGCCCA 701

RESULT 18
US-09-033-556-8
; Sequence 8, Application US/09033556
; Patent No. 6432700
; GENERAL INFORMATION:
; APPLICANT: Henderson, Daniel R.
; TITLE OF INVENTION: YU, De Chao
; TITLE OF INVENTION: ADENOVIRUS VECTORS CONTAINING
; TITLE OF INVENTION: HETEROLOGOUS TRANSCRIPTION REGULATORY ELEMENTS AND METHODS
; TITLE OF INVENTION: OF USING SAME
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,556
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Catherine, Polizzi M
; REGISTRATION NUMBER: 40,130
; REFERENCE/DOCKET NUMBER: 34802-20010.00
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-033-556-8

Query Match 77.0%; Score 15.4; DB 4; Length 858;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GCGCGCCCGCCCGCA 20
DB 20 GCGCGCCCGCCCGCA 36

RESULT 19
US-09-474-699-12
```

```

; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13398
; LENGTH: 807
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13398

Query Match 77.0%; Score 15.4; DB 4; Length 807;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCCA 17
DB 717 CGCGCGCCCGCCCA 701

RESULT 18
US-09-033-556-8
; Sequence 8, Application US/09033556
; Patent No. 6432700
; GENERAL INFORMATION:
; APPLICANT: Henderson, Daniel R.
; TITLE OF INVENTION: YU, De Chao
; TITLE OF INVENTION: ADENOVIRUS VECTORS CONTAINING
; TITLE OF INVENTION: HETEROLOGOUS TRANSCRIPTION REGULATORY ELEMENTS AND METHODS
; TITLE OF INVENTION: OF USING SAME
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,556
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Catherine, Polizzi M
; REGISTRATION NUMBER: 40,130
; REFERENCE/DOCKET NUMBER: 34802-20010.00
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-033-556-8

Query Match 77.0%; Score 15.4; DB 4; Length 858;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GCGCGCCCGCCCGCA 20
DB 20 GCGCGCCCGCCCGCA 36

RESULT 19
US-09-474-699-12
```

; Sequence 12, Application US/09474699  
; Patent No. 6495130  
; GENERAL INFORMATION:  
; APPLICANT: Henderson, Daniel R.  
; APPLICANT: Yu, De Chao  
; TITLE OF INVENTION: TARGET CELL-SPECIFIC ADENOVIRAL VECTORS  
; TITLE OF INVENTION: CONTAINING E3 AND METHODS OF USE THEREOF  
; FILE REFERENCE: 348022001300  
; CURRENT APPLICATION NUMBER: US/09/474,699  
; CURRENT FILING DATE: 1999-12-29  
; PRIOR APPLICATION NUMBER: 60/114,262  
; PRIOR FILING DATE: 1998-12-30  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 858  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-09-474-699-12

Query Match 77.0%; Score 15.4; DB 4; Length 858;  
Best Local Similarity 94.1%; Pred. No. 3.4e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGCGCCCGCCCAAA 20  
DB 20 GGCGCCCGCCCGCAA 36

## RESULT 20

US-09-151-376-8  
; Sequence 8, Application US/09151376  
; Patent No. 6676935  
; GENERAL INFORMATION:  
; APPLICANT: Henderson, D.R.  
; APPLICANT: Schuur, E.R.  
; TITLE OF INVENTION: TISSUE SPECIFIC VIRAL VECTORS  
; FILE REFERENCE: 34802200221  
; CURRENT APPLICATION NUMBER: US/09/151,376  
; CURRENT FILING DATE: 1998-09-10  
; EARLIER APPLICATION NUMBER: 08/669,753  
; EARLIER FILING DATE: 1996-06-26  
; EARLIER APPLICATION NUMBER: 08/495,034  
; EARLIER FILING DATE: 1995-06-27  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 858  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-151-376-8

Query Match 77.0%; Score 15.4; DB 4; Length 858;  
Best Local Similarity 94.1%; Pred. No. 3.4e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGCGCCCGCCCAAA 20  
DB 20 GGCGCCCGCCCGCAA 36

## RESULT 21

US-09-252-991A-7371  
; Sequence 7371, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7371  
; LENGTH: 963  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7371

Query Match 77.0%; Score 15.4; DB 4; Length 963;  
Best Local Similarity 94.1%; Pred. No. 3.4e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCCA 17  
DB 14 CAAGGCGCCCGCCCA 30

## RESULT 22

US-09-252-991A-13494/c  
; Sequence 13494, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13494  
; LENGTH: 984  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13494

Query Match 77.0%; Score 15.4; DB 4; Length 984;  
Best Local Similarity 94.1%; Pred. No. 3.4e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCCA 17  
DB 128 CGCGGCGCCCGCCCA 112

## RESULT 23

US-09-252-991A-3539/c  
; Sequence 3539, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 3539  
; LENGTH: 1002  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-3539

Query Match 77.0%; Score 15.4; DB 4; Length 1002;

```

; Patent No. 6510836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709-2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1997
; LENGTH: 1536
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1997

Query Match 77.0%; Score 15.4; DB 4; Length 1536;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCGAGCCCCAC 18
||| ||||| |||||
DB 653 GAGGCGCCGAGCCCCAC 637

RESULT 27
US-09-252-991A-3548
; Sequence 3548, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3548
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3548

Query Match 77.0%; Score 15.4; DB 4; Length 1644;
Best Local Similarity 94.1%; Pred. No. 3.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCGAGCCCCA 17
||| ||||| |||||
DB 754 CGTGGCGCCGAGCCCCA 770

RESULT 28
US-09-252-991A-1134/c
; Sequence 1134, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```

```
; SEQ ID NO 1134
; LENGTH: 1791
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1134

Query Match      77.0%; Score 15.4; DB 4; Length 1791;
Best Local Similarity 94.1%; Pred. No. 3.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 CGAGGCGCCGAGCCCA 17
Db      736 CGTGGCGCCGAGCCCA 720

; ORGANISM: Homo sapiens
US-09-220-132-188

Query Match      77.0%; Score 15.4; DB 4; Length 2274;
Best Local Similarity 94.1%; Pred. No. 3.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 GAGGCGCCGAGCCCA 18
Db      1584 GAGGCGCCGAGCCCA 1568

Search completed: June 20, 2004, 11:45:36
Job time : 39.2414 secs
```

```
RESULT 29
US-09-301-978C-1
; Sequence 1, Application US/09301978C
; Patent No. 6392015
; GENERAL INFORMATION:
; APPLICANT: Panganiban, Antonito
; APPLICANT: Callahan, Mark A.
; TITLE OF INVENTION: Method of Identifying Modulators of HIV-1 VPU and GAG
; TITLE OF INVENTION: Interaction with U Binding Protein (UBP)
; FILE REFERENCE: 960296.95335
; CURRENT APPLICATION NUMBER: US/09/301,978C
; CURRENT FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,567
; PRIOR FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2221
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1514)
; OTHER INFORMATION: n = any nucleotide.
; NAME/KEY: unsure
; LOCATION: (2066)
; OTHER INFORMATION: n = any nucleotide.
US-09-301-978C-1

Query Match      77.0%; Score 15.4; DB 4; Length 2221;
Best Local Similarity 94.1%; Pred. No. 3.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 GAGGCGCCGAGCCCA 18
Db      660 GAGGCGCCGAGCCCA 676

RESULT 30
US-09-220-132-188/c
; Sequence 188, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shvjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; TITLE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 188
; LENGTH: 2274
; TYPE: DNA
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 113.122 Seconds  
(without alignments)  
809.940 Million cell updates/sec

Title: US-10-624-714-11

Perfect score: 20

Sequence: 1 cgaggccagccacacaa 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	16.8	84.0	993	14	US-10-001-857-91
C 2	16.8	84.0	1002	14	US-10-001-873-6
C 3	16.8	84.0	1094	13	US-10-027-632-9289
C 4	16.8	84.0	1094	13	US-10-027-632-9290
C 5	16.8	84.0	1094	13	US-10-027-632-9291
C 6	16.8	84.0	1094	16	US-10-027-632-9289
C 7	16.8	84.0	1094	16	US-10-027-632-9290
C 8	16.8	84.0	1094	16	US-10-027-632-9291
C 9	16.8	84.0	1094	13	US-10-027-632-9291
C 10	16.8	84.0	9851	15	US-10-017-161-1677
C 11	16.8	84.0	9851	15	US-10-292-798-1337
C 12	16.8	84.0	21501	15	US-10-006-430-11
C 13	16.8	84.0	28000	15	US-10-091-625-11
C 14	16.8	84.0	28000	15	US-10-096-399A-11
C 15	16.8	84.0	28000	15	US-10-096-399A-11

Sequence 11, Appl	US-10-461-668-11	16	28000	84.0	16.8
Sequence 381, App	US-10-388-263-381	16	28000	84.0	16.8
Sequence 11, Appl	US-09-999-121-11	10	37113	84.0	16.8
Sequence 1789, Ap	US-09-814-353-1789	10	475	82.0	16.4
Sequence 8137, Ap	US-09-814-353-8137	10	475	82.0	16.4
Sequence 1521, A	US-09-814-353-14521	10	565	82.0	16.4
Sequence 47, Appl	US-10-275-595A-47	17	3287	82.0	16.4
Sequence 308, App	US-10-108-260A-308	16	3370	82.0	16.4
Sequence 2078, A	US-09-814-353-20078	10	3592	82.0	16.4
Sequence 226, App	US-10-085-117-226	16	246144	82.0	16.4
Sequence 754, App	US-10-087-132-754	13	276276	82.0	16.4
Sequence 11, App	US-10-211-462-112	13	3845	80.0	16
Sequence 11, App	US-10-007-926A-110	15	3845	80.0	16
Sequence 34, Appl	US-10-101-510-4	15	3845	80.0	16
Sequence 4, Appl	US-10-021-660-34	15	3845	80.0	16
Sequence 162, App	US-10-116-802-162	13	3914	80.0	16
Sequence 148, App	US-10-044-090-148	14	3914	80.0	16
Sequence 69, Appl	US-09-373-658-69	10	8670	80.0	16
Sequence 69, Appl	US-09-993-687-69	11	8670	80.0	16
Sequence 284, App	US-09-873-367C-284	13	149480	80.0	16
Sequence 285, App	US-09-873-367C-285	10	149480	80.0	16
Sequence 400, App	US-10-087-192-400	13	218155	80.0	16
Sequence 40951, A	US-10-424-599-40951	13	245	79.0	15.8
Sequence 11265, A	US-10-085-783A-11265	13	268	79.0	15.8
Sequence 11265, A	US-10-243-535A-11265	16	268	79.0	15.8
Sequence 83608, A	US-10-424-599-83608	13	340	79.0	15.8
Sequence 58186, A	US-10-424-599-58186	13	341	79.0	15.8
Sequence 28160, A	US-09-918-995-28160	10	461	79.0	15.8
Sequence 13759, A	US-09-918-995-13759	10	492	79.0	15.8
Sequence 1, Appl	US-09-931-186-1	9	786	79.0	15.8
Sequence 3, Appl	US-09-931-186-3	9	786	79.0	15.8
Sequence 5, Appl	US-09-931-186-5	9	786	79.0	15.8
Sequence 7, Appl	US-09-931-186-7	9	786	79.0	15.8
Sequence 19, Appl	US-09-931-186-19	9	786	79.0	15.8
Sequence 22, Appl	US-09-931-186-22	9	786	79.0	15.8
Sequence 142002, A	US-10-027-632-142002	13	874	79.0	15.8
Sequence 163596, A	US-10-027-632-163596	16	874	79.0	15.8
Sequence 142002, A	US-10-027-632-142002	16	874	79.0	15.8
Sequence 163596, A	US-10-027-632-163596	16	874	79.0	15.8
Sequence 240, App	US-09-962-832-240	9	947	79.0	15.8
Sequence 3418, App	US-09-880-107-3418	9	947	79.0	15.8
Sequence 324, App	US-10-332-859-324	17	1004	79.0	15.8
Sequence 18, Appl	US-09-984-245-18	10	1054	79.0	15.8
Sequence 18, Appl	US-09-966-262-18	10	1054	79.0	15.8
Sequence 18, Appl	US-09-983-966-18	10	1054	79.0	15.8
Sequence 18, Appl	US-10-059-395-18	13	1054	79.0	15.8
Sequence 18, Appl	US-10-143-090-18	15	1054	79.0	15.8
Sequence 684, App	US-09-925-300-684	9	1251	79.0	15.8
Sequence 485, App	US-10-296-115-485	13	1256	79.0	15.8
Sequence 99825, A	US-10-027-632-99825	16	2042	79.0	15.8
Sequence 99825, A	US-10-027-632-99825	16	2042	79.0	15.8
Sequence 1695, Ap	US-10-108-260A-1695	16	2042	79.0	15.8
Sequence 284, App	US-10-104-047-284	16	2731	79.0	15.8
Sequence 8596, Ap	US-09-764-891-8596	9	2763	79.0	15.8
Sequence 1331, Ap	US-09-764-847-1331	10	7755	79.0	15.8
Sequence 8599, Ap	US-09-764-891-8599	10	7755	79.0	15.8
Sequence 1331, Ap	US-10-093-154-1331	15	7755	79.0	15.8
Sequence 582, App	US-10-296-115-582	13	8029	79.0	15.8
Sequence 47, Appl	US-10-007-926A-47	15	13268	79.0	15.8
Sequence 75, Appl	US-10-037-270-75	15	13857	79.0	15.8
Sequence 75, Appl	US-10-117-722-75	16	13857	79.0	15.8
Sequence 648, App	US-10-080-170-648	16	86114	79.0	15.8
Sequence 1470, Ap	US-10-240-425-1470	17	100267	79.0	15.8
Sequence 68, Appl	US-10-717-597-68	17	126990	79.0	15.8
Sequence 4, Appl	US-10-467-164-4	17	456	77.0	15.4
Sequence 7509, Ap	US-10-282-122A-7509	13	609	77.0	15.4
Sequence 350, App	US-10-389-647-350	13	609	77.0	15.4
Sequence 140280, A	US-10-027-632-140280	13	666	77.0	15.4
Sequence 140280, A	US-10-027-632-140280	16	666	77.0	15.4
Sequence 117405, A	US-10-027-632-117405	13	734	77.0	15.4
Sequence 6, Appl	US-09-392-822-6	9	858	77.0	15.4
Sequence 8, Appl	US-09-151-376-8	10	858	77.0	15.4

88	15.4	77.0	858	10	US-09-814-357-15	Sequence 15, Appl	C 161	15.2	76.0	632	16	US-10-027-632-55223	Sequence 55223, A
89	15.4	77.0	858	10	US-09-814-351-15	Sequence 15, Appl	C 162	15.2	76.0	632	16	US-10-027-632-55224	Sequence 55224, A
90	15.4	77.0	858	10	US-10-226-820-12	Sequence 18, Appl	C 163	15.2	76.0	632	16	US-10-027-632-314381	Sequence 314381, A
91	15.4	77.0	858	10	US-10-139-089-8	Sequence 8, Appl	C 164	15.2	76.0	632	16	US-10-027-632-314382	Sequence 314382, A
92	15.4	77.0	884	13	US-10-343-650A-93	Sequence 93, Appl	C 165	15.2	76.0	634	10	US-09-814-353-5055	Sequence 5055, Ap
93	15.4	77.0	884	13	US-10-343-650A-93	Sequence 21, Appl	C 166	15.2	76.0	634	10	US-09-814-353-11347	Sequence 11347, A
94	15.4	77.0	900	15	US-10-257-080-21	Sequence 2, Appl	C 167	15.2	76.0	665	12	US-10-152-319A-234	Sequence 234, App
95	15.4	77.0	900	17	US-10-467-184-1	Sequence 1, Appl	C 168	15.2	76.0	665	12	US-10-152-319A-234	Sequence 166907, A
96	15.4	77.0	993	9	US-09-911-005-3	Sequence 3, Appl	C 169	15.2	76.0	688	13	US-10-027-632-166907	Sequence 166907, A
97	15.4	77.0	993	13	US-10-343-650A-73	Sequence 73, Appl	C 170	15.2	76.0	688	13	US-10-027-632-166907	Sequence 22463, A
98	15.4	77.0	993	15	US-10-145-586-32	Sequence 32, Appl	C 171	15.2	76.0	754	13	US-10-027-632-22463	Sequence 22463, A
99	15.4	77.0	993	17	US-10-432-101-1	Sequence 1, Appl	C 172	15.2	76.0	754	13	US-10-027-632-22463	Sequence 4674, Ap
100	15.4	77.0	1038	15	US-10-225-632A-671	Sequence 671, App	C 173	15.2	76.0	795	15	US-10-156-761-4674	Sequence 28576, A
101	15.4	77.0	1156	13	US-10-027-632-9071	Sequence 9071, Ap	C 174	15.2	76.0	906	13	US-10-282-132A-28576	Sequence 27676, A
102	15.4	77.0	1156	16	US-10-027-632-9071	Sequence 9071, Ap	C 175	15.2	76.0	906	13	US-10-282-132A-28576	Sequence 27, Appl
103	15.4	77.0	1361	15	US-10-188-405-3	Sequence 3, Appl	C 176	15.2	76.0	1126	15	US-10-000-897-27	Sequence 42803, A
104	15.4	77.0	1393	15	US-10-017-161-1715	Sequence 1715, Ap	C 177	15.2	76.0	1382	16	US-10-369-493-42803	Sequence 317, App
105	15.4	77.0	1393	16	US-10-282-798-1371	Sequence 1371, Ap	C 178	15.2	76.0	1405	15	US-10-006-285-317	Sequence 42934, A
106	15.4	77.0	1396	9	US-09-911-005-1	Sequence 1, Appl	C 179	15.2	76.0	1806	15	US-10-168-097A-1	Sequence 1, Appl
107	15.4	77.0	1396	15	US-10-145-586-30	Sequence 30, Appl	C 180	15.2	76.0	1806	15	US-10-159-563-164	Sequence 164, App
108	15.4	77.0	1423	13	US-10-302-172-886	Sequence 886, App	C 181	15.2	76.0	2420	16	US-10-108-260A-418	Sequence 418, App
109	15.4	77.0	1847	17	US-10-333-946-23	Sequence 23, Appl	C 182	15.2	76.0	2824	13	US-10-282-122A-25564	Sequence 25564, A
110	15.4	77.0	1992	15	US-10-156-761-5203	Sequence 5203, Ap	C 183	15.2	76.0	2850	15	US-10-156-761-364	Sequence 364, App
111	15.4	77.0	2116	16	US-10-108-260A-1493	Sequence 1493, Ap	C 184	15.2	76.0	3183	15	US-10-000-897-29	Sequence 29, Appl
112	15.4	77.0	2221	14	US-10-090-378-1	Sequence 1, Appl	C 185	15.2	76.0	3215	15	US-10-000-897-28	Sequence 28, Appl
113	15.4	77.0	2294	9	US-09-833-790-423	Sequence 423, App	C 186	15.2	76.0	3246	13	US-10-280-576-23	Sequence 23, Appl
114	15.4	77.0	2341	15	US-10-181-612-1	Sequence 1, Appl	C 187	15.2	76.0	3398	13	US-10-302-172-802	Sequence 802, App
115	15.4	77.0	2341	15	US-10-181-612-7	Sequence 7, Appl	C 188	15.2	76.0	3398	13	US-10-342-887-20	Sequence 20, Appl
116	15.4	77.0	2615	13	US-10-424-589-98605	Sequence 98605, A	C 189	15.2	76.0	3358	13	US-10-172-118-20	Sequence 20, Appl
117	15.4	77.0	2941	13	US-10-247-703-25	Sequence 25, Appl	C 190	15.2	76.0	4146	13	US-10-282-122A-28711	Sequence 28711, A
118	15.4	77.0	3113	16	US-10-062-674-1572	Sequence 1572, Ap	C 191	15.2	76.0	4545	16	US-10-259-194A-145	Sequence 145, App
119	15.4	77.0	4657	9	US-09-962-436-267	Sequence 267, App	C 192	15.2	76.0	14061	16	US-10-093-463-73	Sequence 73, Appl
120	15.4	77.0	8181	16	US-10-447-839A-18	Sequence 18, Appl	C 193	15.2	76.0	14109	16	US-10-093-463-71	Sequence 71, Appl
121	15.4	77.0	8186	13	US-10-247-703-23	Sequence 23, Appl	C 194	15.2	76.0	22567	13	US-10-087-192-979	Sequence 979, App
122	15.4	77.0	8186	15	US-10-029-517-19	Sequence 19, Appl	C 195	15.2	76.0	23907	14	US-10-077-130-6	Sequence 6, Appl
123	15.4	77.0	13635	13	US-10-210-175-13	Sequence 13, Appl	C 196	15.2	76.0	23909	13	US-10-087-192-604	Sequence 604, App
124	15.4	77.0	13635	17	US-10-464-368-30	Sequence 30, Appl	C 197	15.2	76.0	24120	14	US-10-077-130-4	Sequence 4, Appl
125	15.4	77.0	13836	13	US-10-087-192-671	Sequence 671, App	C 198	15.2	76.0	49600	17	US-10-459-262A-2	Sequence 2, Appl
126	15.4	77.0	14887	13	US-10-276-774-373	Sequence 373, App	C 199	15.2	76.0	63294	12	US-09-997-722-205	Sequence 205, App
127	15.4	77.0	14889	15	US-10-101-510-356	Sequence 356, App	C 200	15.2	76.0	91552	13	US-10-415-058-5	Sequence 5, Appl
128	15.4	77.0	14896	12	US-09-750-972-6	Sequence 6, Appl	C 201	15.2	76.0	172369	13	US-10-087-192-1366	Sequence 1366, Ap
129	15.4	77.0	14896	16	US-10-159-563-206	Sequence 206, App	C 202	15.2	76.0	172570	17	US-10-450-826-114	Sequence 114, App
130	15.4	77.0	15202	15	US-10-420-845-21	Sequence 21, Appl	C 203	15.2	76.0	344548	13	US-10-087-192-334	Sequence 334, App
131	15.4	77.0	21799	16	US-10-074-024-724	Sequence 404, App	C 204	15.2	76.0	9025608	15	US-10-156-761-1	Sequence 1, Appl
132	15.4	77.0	24707	9	US-09-740-027-3	Sequence 3, Appl	C 205	15	75.0	407	9	US-09-960-352-3409	Sequence 3409, Ap
133	15.4	77.0	24707	15	US-10-274-968-3	Sequence 3, Appl	C 206	15	75.0	419	10	US-09-918-995-8599	Sequence 8599, Ap
134	15.4	77.0	30346	15	US-10-004-113-28	Sequence 28, Appl	C 207	15	75.0	431	12	US-09-876-143-881	Sequence 881, App
135	15.4	77.0	30346	17	US-10-394-948-16	Sequence 16, Appl	C 208	15	75.0	534	13	US-10-027-632-242774	Sequence 242774, A
136	15.4	77.0	104083	13	US-10-087-192-670	Sequence 670, App	C 209	15	75.0	534	16	US-10-027-632-242774	Sequence 242774, A
137	15.4	77.0	129722	17	US-10-450-826-24	Sequence 24, Appl	C 210	15	75.0	576	13	US-10-027-632-257072	Sequence 257072, A
138	15.4	77.0	9025608	15	US-10-156-761-1	Sequence 1, Appl	C 211	15	75.0	576	16	US-10-027-632-257072	Sequence 257072, A
139	15.2	76.0	353	9	US-09-867-701-4104	Sequence 4104, App	C 212	15	75.0	759	13	US-10-282-122A-29944	Sequence 29944, A
140	15.2	76.0	412	9	US-09-833-381-649	Sequence 649, App	C 213	15	75.0	936	13	US-10-282-122A-19970	Sequence 19970, A
141	15.2	76.0	414	10	US-09-918-995-16966	Sequence 16966, A	C 214	15	75.0	1129	13	US-10-027-632-259164	Sequence 259164, A
142	15.2	76.0	437	9	US-09-867-701-4258	Sequence 4258, Ap	C 215	15	75.0	1129	13	US-10-027-632-259165	Sequence 259165, A
143	15.2	76.0	439	9	US-09-960-352-12527	Sequence 12527, A	C 216	15	75.0	1129	16	US-10-027-632-259164	Sequence 259164, A
144	15.2	76.0	520	13	US-10-027-632-199028	Sequence 199028, A	C 217	15	75.0	1129	16	US-10-027-632-259165	Sequence 259165, A
145	15.2	76.0	520	16	US-10-027-632-199028	Sequence 199028, A	C 218	15	75.0	1146	13	US-10-027-632-200286	Sequence 200286, A
146	15.2	76.0	621	16	US-10-260-238-187	Sequence 187, App	C 219	15	75.0	1146	16	US-10-027-632-200286	Sequence 200286, A
147	15.2	76.0	627	13	US-10-027-632-393	Sequence 393, App	C 220	15	75.0	35169	13	US-10-087-192-91	Sequence 91, Appl
148	15.2	76.0	627	13	US-10-027-632-394	Sequence 394, App	C 221	15	75.0	35169	13	US-10-087-192-1624	Sequence 1624, Ap
149	15.2	76.0	627	16	US-10-027-632-393	Sequence 393, App	C 222	15	75.0	126990	17	US-10-717-597-68	Sequence 68, Appl
150	15.2	76.0	627	16	US-10-027-632-394	Sequence 394, App	C 223	15	75.0	163321	13	US-10-087-192-76	Sequence 76, Appl
151	15.2	76.0	628	13	US-10-424-599-14673	Sequence 14673, A	C 224	14.8	74.0	160	9	US-09-864-761-21543	Sequence 21543, A
152	15.2	76.0	629	10	US-09-814-353-17731	Sequence 17731, A	C 225	14.8	74.0	240	9	US-09-764-869-1590	Sequence 1590, Ap
153	15.2	76.0	632	13	US-10-027-632-54636	Sequence 54636, A	C 226	14.8	74.0	240	15	US-10-091-504-1590	Sequence 1590, Ap
154	15.2	76.0	632	13	US-10-027-632-54637	Sequence 54637, A	C 227	14.8	74.0	240	16	US-10-227-577-1590	Sequence 1590, Ap
155	15.2	76.0	632	13	US-10-027-632-55223	Sequence 55223, A	C 228	14.8	74.0	247	15	US-10-116-265-3	Sequence 3, Appl
156	15.2	76.0	632	13	US-10-027-632-55224	Sequence 55224, A	C 229	14.8	74.0	253	9	US-09-764-869-1587	Sequence 1587, Ap
157	15.2	76.0	632	13	US-10-027-632-314381	Sequence 314381, A	C 230	14.8	74.0	253	9	US-09-764-869-1588	Sequence 1588, Ap
158	15.2	76.0	632	13	US-10-027-632-314382	Sequence 314382, A	C 231	14.8	74.0	253	15	US-10-091-504-1587	Sequence 1587, Ap
159	15.2	76.0	632	13	US-10-027-632-54636	Sequence 54636, A	C 232	14.8	74.0	253	15	US-10-091-504-1588	Sequence 1588, Ap
160	15.2	76.0	632	16	US-10-027-632-54637	Sequence 54637, A	C 233	14.8	74.0	253	16	US-10-227-577-1590	Sequence 1590, Ap



APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9289  
; LENGTH: 1094  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-9289

Query Match 84.0%; Score 16.8; DB 13; Length 1094;  
Best Local Similarity 90.0%; Pred. No. 2.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCAACAA 20  
Db 170 CGATGCGTCCAGCCCAACAA 189

## RESULT 4

US-10-027-632-9290  
; Sequence 9290, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:

APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9290  
; LENGTH: 1094  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-9290

Query Match 84.0%; Score 16.8; DB 13; Length 1094;  
Best Local Similarity 90.0%; Pred. No. 2.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCAACAA 20  
Db 170 CGATGCGTCCAGCCCAACAA 189

## RESULT 5

US-10-027-632-9291  
; Sequence 9291, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9291  
; LENGTH: 1094  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-9291

Query Match 84.0%; Score 16.8; DB 13; Length 1094;  
Best Local Similarity 90.0%; Pred. No. 2.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCAACAA 20  
Db 170 CGATGCGTCCAGCCCAACAA 189

## RESULT 6

US-10-027-632-9289  
; Sequence 9289, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09

```
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9289
; LENGTH: 1094
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-9289

Query Match      84.0%; Score 16.8; DB 16; Length 1094;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1 CGAGGCGCCAGCCCAAA 20
      |||||
Db      170 CGATGCGTCCAGCCCAAA 189

RESULT 7
US-10-027-632-9290
; Sequence 9290, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9290
; LENGTH: 1094
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-9290

Query Match      84.0%; Score 16.8; DB 16; Length 1094;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1 CGAGGCGCCAGCCCAAA 20
      |||||
Db      170 CGATGCGTCCAGCCCAAA 189

RESULT 8
US-10-027-632-9291
; Sequence 9291, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9290
; LENGTH: 1094
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-9290

Query Match      84.0%; Score 16.8; DB 16; Length 1094;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1 CGAGGCGCCAGCCCAAA 20
      |||||
Db      170 CGATGCGTCCAGCCCAAA 189

RESULT 9
US-10-282-122A-40410/c
; Sequence 40410, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40410
```

```
/ LENGTH: 1830
/ TYPE: DNA
/ ORGANISM: Treponema pallidum
US-10-282-122A-40410

Query Match      84.0%; Score 16.8; DB 13; Length 1830;
Best Local Similarity 90.0%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGAGGGCCCGCCACCA 20
Db      660 CGAGGGCCCAATCCCA 641

RESULT 10
US-10-017-161-1677
/ Sequence 1677, Application US/10017161
/ Publication No. US2003014368A1
/ GENERAL INFORMATION:
/ APPLICANT: SUWA, MAKIKO
/ APPLICANT: ASAI, KIYOSHI
/ APPLICANT: AKIYAMA, YUTAKA
/ APPLICANT: ABURATANI, HIROYUKI
/ TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
/ FILE REFERENCE: 084335/0152
/ CURRENT APPLICATION NUMBER: US/10/017,161
/ CURRENT FILING DATE: 2002-12-18
/ PRIOR APPLICATION NUMBER: JP 2001/246789
/ PRIOR FILING DATE: 2001-06-18
/ NUMBER OF SEQ ID NOS: 2430
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1677
/ LENGTH: 9851
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: source
/ LOCATION: (1)..(9851)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (201)..(581)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (2244)..(2394)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (2475)..(3642)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (4153)..(4238)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (4840)..(4953)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (5064)..(5179)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (5444)..(5833)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (6503)..(7011)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (7503)..(7674)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (7925)..(8081)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (8179)..(8325)
/ FEATURE:
/ NAME/KEY: CDS

/ LOCATION: (8393)..(8945)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (9525)..(9651)
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1066)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1068)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1094)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1097)..(1098)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1100)..(1101)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1126)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1129)..(1133)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1135)..(1139)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1141)..(1143)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1189)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1191)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1196)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1213)..(1214)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1245)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1250)..(1251)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1260)..(1269)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (1271)..(1274)
/ OTHER INFORMATION: a, t, c, g, unknown or other
/ FEATURE:
```



```

; LOCATION: (1250)..(1251)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1260)..(1269)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1271)..(1274)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1278)..(1377)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1438)..(1438)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1441)..(1441)
; OTHER INFORMATION: a, t, c, g, unknown or other
; US-10-292-798-1337

```

```

Query Match      84.0%; Score 16.8; DB 16; Length 9851;
Best Local Similarity 90.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 CGAGGCGCCAGCCACAA 20
Db      8208 CCAGCGCCAGCCACAA 8227

```

## RESULT 12

```

US-10-006-430-11/c
; Sequence 11, Application US/10006430
; Publication No. US20030113914A1
; GENERAL INFORMATION:
; APPLICANT: Mark J. Graham
; TITLE OF INVENTION: ANTISENSE MODULATION OF CD81 EXPRESSION
; FILE REFERENCE: RTS-0341
; CURRENT APPLICATION NUMBER: US/10/006,430
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 11
; LENGTH: 21501
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (518)...(812)
; OTHER INFORMATION: exon 1
; NAME/KEY: exon:intron junction
; LOCATION: (812)...(813)
; OTHER INFORMATION: exon 1:intron 1
; NAME/KEY: intron
; LOCATION: (813)...(13608)
; OTHER INFORMATION: intron 1
; NAME/KEY: intron:exon junction
; LOCATION: (13608)...(13609)
; OTHER INFORMATION: intron 1:exon 2
; NAME/KEY: exon
; LOCATION: (13609)...(13723)
; OTHER INFORMATION: exon 2
; NAME/KEY: exon:intron junction
; LOCATION: (13723)...(13724)
; OTHER INFORMATION: exon 2:intron 2
; NAME/KEY: intron
; LOCATION: (13724)...(17291)
; OTHER INFORMATION: intron 2
; NAME/KEY: intron:exon junction
; LOCATION: (17291)...(17292)

```

```

; OTHER INFORMATION: intron 2:exon 3
; NAME/KEY: exon
; LOCATION: (17292)...(17389)
; OTHER INFORMATION: exon 3
; NAME/KEY: exon:intron junction
; LOCATION: (17389)...(17390)
; OTHER INFORMATION: exon 3:intron 3
; NAME/KEY: intron
; LOCATION: (17390)...(18169)
; OTHER INFORMATION: intron 3
; NAME/KEY: intron:exon junction
; LOCATION: (18169)...(18169)
; OTHER INFORMATION: intron 3:exon 4
; NAME/KEY: exon
; LOCATION: (18169)...(18243)
; OTHER INFORMATION: exon 4
; NAME/KEY: exon:intron junction
; LOCATION: (18243)...(18244)
; OTHER INFORMATION: exon 4:intron 4
; NAME/KEY: intron
; LOCATION: (18244)...(18612)
; OTHER INFORMATION: intron 4
; NAME/KEY: intron:exon junction
; LOCATION: (18612)...(18613)
; OTHER INFORMATION: intron 4:exon 5
; NAME/KEY: exon
; LOCATION: (18613)...(18717)
; OTHER INFORMATION: exon 5
; NAME/KEY: exon:intron junction
; LOCATION: (18717)...(18718)
; OTHER INFORMATION: exon 5:intron 5
; NAME/KEY: intron
; LOCATION: (18718)...(19065)
; OTHER INFORMATION: intron 5
; NAME/KEY: intron:exon junction
; LOCATION: (19065)...(19066)
; OTHER INFORMATION: intron 5:exon 6
; NAME/KEY: exon
; LOCATION: (19066)...(19167)
; OTHER INFORMATION: exon 6
; NAME/KEY: exon:intron junction
; LOCATION: (19167)...(19168)
; OTHER INFORMATION: exon 6:intron 6
; NAME/KEY: intron
; LOCATION: (19168)...(19824)
; OTHER INFORMATION: intron 6
; NAME/KEY: intron:exon junction
; LOCATION: (19824)...(19825)
; OTHER INFORMATION: intron 6:exon 7
; NAME/KEY: exon
; LOCATION: (19825)...(19911)
; OTHER INFORMATION: exon 7
; NAME/KEY: exon:intron junction
; LOCATION: (19911)...(19912)
; OTHER INFORMATION: exon 7:intron 7
; NAME/KEY: intron
; LOCATION: (19912)...(20000)
; OTHER INFORMATION: intron 7
; NAME/KEY: intron:exon junction
; LOCATION: (20000)...(20001)
; OTHER INFORMATION: intron 7:exon 8
; NAME/KEY: exon
; LOCATION: (20001)...(20610)
; OTHER INFORMATION: exon 8
; US-10-006-430-11

```

```

Query Match      84.0%; Score 16.8; DB 15; Length 21501;
Best Local Similarity 90.0%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 CGAGGCGCCAGCCACAA 20
Db      7509 CAAGGCGCCAGCCACAA 7490

```



```
RESULT 13
US-10-091-625-11
; Sequence 11, Application US/10091625
; Publication No. US20030170636A1
; GENERAL INFORMATION:
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF JAGGED 2 EXPRESSION
; FILE REFERENCE: RTS-0244
; CURRENT APPLICATION NUMBER: US/10/091,625
; CURRENT FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 11
; LENGTH: 28000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-10-091-625-11
Query Match      84.0%; Score 16.8; DB 15; Length 28000;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGAGCGCCCGCCAGCCCAAA 20
Db      26057 CCAGCGCCCGCCAGCCCAAA 26076

RESULT 14
US-10-096-399A-11
; Sequence 11, Application US/10096399A
; Publication No. US20030185829A1
; GENERAL INFORMATION:
; APPLICANT: Koller, Erich
; APPLICANT: Shepherd, Peter J.
; TITLE OF INVENTION: JAGGED 2 INHIBITORS FOR INDUCING APOPTOSIS
; FILE REFERENCE: ISPH-0660
; CURRENT APPLICATION NUMBER: US/10/096,399A
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 28000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-096-399A-11
Query Match      84.0%; Score 16.8; DB 15; Length 28000;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGAGCGCCCGCCAGCCCAAA 20
Db      26057 CCAGCGCCCGCCAGCCCAAA 26076

RESULT 15
US-10-461-668-11
; Sequence 11, Application US/10461668
; Publication No. US20030207839A1
; GENERAL INFORMATION:
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF JAGGED 2 EXPRESSION
; FILE REFERENCE: RTS-0244
; CURRENT APPLICATION NUMBER: US/10/461,668
; CURRENT FILING DATE: 2003-05-13
; PRIOR APPLICATION NUMBER: US/10/091,625
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 11
; LENGTH: 28000
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
; FEATURE:
US-10-461-668-11
Query Match      84.0%; Score 16.8; DB 16; Length 28000;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGAGCGCCCGCCAGCCCAAA 20
Db      26057 CCAGCGCCCGCCAGCCCAAA 26076

RESULT 16
US-10-388-263-381
; Sequence 381, Application US/10388263
; Publication No. US20030228597A1
; GENERAL INFORMATION:
; APPLICANT: Cosset, Lex M.
; APPLICANT: Baker, Brenda F.
; APPLICANT: McNeill, John
; APPLICANT: Freier, Susan M.
; APPLICANT: Sasnor, Henri M.
; APPLICANT: Brooks, Douglas G.
; APPLICANT: Ohashi, Cara
; APPLICANT: Wyatt, Jacqueline R.
; APPLICANT: Borchers, Alexander
; APPLICANT: Vickers, Timothy A.
; TITLE OF INVENTION: IDENTIFICATION OF GENETIC TARGETS FOR
; TITLE OF INVENTION: MODULATION BY OLIGONUCLEOTIDES AND
; TITLE OF INVENTION: GENERATION OF OLIGONUCLEOTIDES FOR GENE MODULATION
; FILE REFERENCE: ISIS-4503
; CURRENT APPLICATION NUMBER: US/10/388,263
; CURRENT FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 947
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 381
; LENGTH: 28000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-388-263-381
Query Match      84.0%; Score 16.8; DB 16; Length 28000;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGAGCGCCCGCCAGCCCAAA 20
Db      26057 CCAGCGCCCGCCAGCCCAAA 26076

RESULT 17
US-09-999-121-11/c
; Sequence 11, Application US/09999121
; Publication No. US20030039982A1
; GENERAL INFORMATION:
; APPLICANT: Ryan, James W.
; TITLE OF INVENTION: ISOLATED GENOMIC NUCLEOTIDE FRAGMENTS FROM THE p15 REGION OF
; FILE REFERENCE: JR-10,001-US
; CURRENT APPLICATION NUMBER: US/09/999,121
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/244,705
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 37113
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-999-121-11
Query Match      84.0%; Score 16.8; DB 10; Length 37113;
```

```

; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8137
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 111
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-8137

Query Match 82.0%; Score 16.4; DB 10; Length 475;
Best Local Similarity 94.4%; Pred. No. 3.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGCGCCCCAGCCCCAC 18
DB 286 CGAGCGCCCCAGCCCCAC 303

RESULT 20
US-09-814-353-14521
; Sequence 14521, Application US/09814953
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14521
; LENGTH: 565
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-14521

Query Match 82.0%; Score 16.4; DB 10; Length 565;
Best Local Similarity 94.4%; Pred. No. 3.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGCGCCCCAGCCCCAC 18
DB 288 CGAGCGCCCCAGCCCCAC 305

```

US-10-275-595A-47  
; Sequence 47, Application US/10275595A  
; Publication No. US20040078804A1  
; GENERAL INFORMATION:  
; APPLICANT: YUE, Henry  
; APPLICANT: TANG, Y. Tom  
; APPLICANT: AU-YOUNG, Janice  
; APPLICANT: LU, Dyung Aina M.  
; APPLICANT: BAUGHN, Maria R.  
; APPLICANT: HILLMAN, Jennifer L.  
; APPLICANT: AZIMZAI, Valda  
; APPLICANT: LAL, Preeti  
; APPLICANT: YAO, Monique G.  
; APPLICANT: BANDMAN, Olga  
; APPLICANT: BURFORD, Neil  
; APPLICANT: BATRA, Sajeev  
; APPLICANT: KEARNEY, Liam  
; APPLICANT: POLICKY, Jennifer L.  
; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS  
; FILE REFERENCE: PF-0772 USN  
; CURRENT APPLICATION NUMBER: US/10/275,595A  
; CURRENT FILING DATE: 2003-08-13  
; PRIOR APPLICATION NUMBER: US 60/201,960  
; PRIOR FILING DATE: 2000-05-05  
; PRIOR APPLICATION NUMBER: US 60/202,729  
; PRIOR FILING DATE: 2000-05-08  
; PRIOR APPLICATION NUMBER: US 60/209,705  
; PRIOR FILING DATE: 2000-06-05  
; PRIOR APPLICATION NUMBER: US 60/210,149  
; PRIOR FILING DATE: 2000-06-07  
; PRIOR APPLICATION NUMBER: US 60/213,215  
; PRIOR FILING DATE: 2000-06-21  
; NUMBER OF SEQ ID NOS: 68  
; SOFTWARE: PERL Program  
; SEQ ID NO 47  
; LENGTH: 3287  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No: 1295420CE1  
US-10-275-595A-47

Query Match 82.0%; Score 16.4; DB 17; Length 3287;  
Best Local Similarity 94.4%; Pred. No. 2.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCCGCCAC 18  
Db 220 CGAGGCGCCCGCCAC 237

RESULT 22  
US-10-108-260A-308  
; Sequence 308, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US20040005560A1e1 full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 308  
; LENGTH: 3370  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-108-260A-308

Query Match 82.0%; Score 16.4; DB 16; Length 3370;  
Best Local Similarity 94.4%; Pred. No. 2.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCCGCCAC 18  
Db 312 CGAGGCGCCCGCCAC 329

RESULT 23  
US-09-814-353-20078  
; Sequence 20078, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20078  
; LENGTH: 3592  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 1, 2, 3, 4, 3590, 3591, 3592  
; OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-20078

Query Match 82.0%; Score 16.4; DB 10; Length 3592;  
Best Local Similarity 94.4%; Pred. No. 2.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCCGCCAC 18  
Db 292 CGAGGCGCCCGCCAC 309

RESULT 24  
US-10-085-117-226/c  
; Sequence 226, Application US/10085117  
; Publication No. US2003023334A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
; FILE REFERENCE: 529452000121  
; CURRENT APPLICATION NUMBER: US/10/085,117  
; CURRENT FILING DATE: 2002-02-27  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 361  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 226  
; LENGTH: 246144  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:

```
; NAME/KEY: variation
; LOCATION: (1)....(246144)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-226

Query Match      82.0%; Score 16.4; DB 15; Length 246144;
Best Local Similarity 94.4%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 GAGGCGCCAGCCGCCACA 19
Db      14025 GAGGCGCTAGCCGCCACA 14008

RESULT 25
US-10-087-192-754/c
; Sequence 754, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: MORRIS, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER.
; FILE REFERENCE: 52945200122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 754
; LENGTH: 276276
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-087-192-754

Query Match      82.0%; Score 16.4; DB 13; Length 276276;
Best Local Similarity 94.4%; Pred. No. 93;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 CGAGGCGCCAGCCGCCAC 18
Db      133204 CGAGGCGCCAGCCGCCAC 133187

RESULT 26
US-10-211-462-112/c
; Sequence 112, Application US/10211462
; Publication No. US20040033495A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Aziz, Natasha
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and
; TITLE OF INVENTION: Methods of Screening for Angiogenesis Modulators
; FILE REFERENCE: 018501-006200US
; CURRENT APPLICATION NUMBER: US/10/211,462
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: US 09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/791,390
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 60/310,025
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/334,244
; PRIOR FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 112

; NAME/KEY: variation
; LOCATION: (1)....(3845)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-226

Query Match      80.0%; Score 16; DB 13; Length 3845;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 GCGCCCGAGCCGCCACA 20
Db      693 GCGCCCGAGCCGCCACA 678

RESULT 27
US-10-007-926A-110/c
; Sequence 110, Application US/10007926A
; Publication No. US20030143539A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, FRANCOIS
; APPLICANT: HOULGATTE, REMI
; APPLICANT: BIRNBAUM, DANIEL
; APPLICANT: NGUYEN, CATHERINE
; APPLICANT: VIENS, PATRICE
; APPLICANT: FERT, VINCENT
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS
; TITLE OF INVENTION: USING ARRAYS OF CANDIDATE GENES
; FILE REFERENCE: 1546-R-00
; CURRENT APPLICATION NUMBER: US/10/007,926A
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,090
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 110
; LENGTH: 3845
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: tyrosine kinase with immunoglobulin and
; OTHER INFORMATION: epidermal growth factor homology domains (TIE)
; OTHER INFORMATION: Gene.
US-10-007-926A-110

Query Match      80.0%; Score 16; DB 15; Length 3845;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 GCGCCCGAGCCGCCACA 20
Db      693 GCGCCCGAGCCGCCACA 678

RESULT 28
US-10-101-510-4/c
; Sequence 4, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117, 0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3845
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-4
```

Query Match 80.0%; Score 16; DB 15; Length 3845;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGCCAGCCCCACAA 20  
|||  
Db 693 GCGCCAGCCCCACAA 678

## RESULT 29

US-10-021-660-34/c  
; Sequence 34, Application US/10021660  
; Publication No. US20030152926A1  
; GENERAL INFORMATION:  
; APPLICANT: Murray, Richard  
; APPLICANT: Glynn, Richard  
; APPLICANT: Watson, Susan R.  
; APPLICANT: EOS Biotechnology, Inc.  
; TITLE OF INVENTION: No. US20030152926A1 Methods of Diagnosis of Angiogenesis,  
; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis  
; TITLE OF INVENTION: Modulators  
; FILE REFERENCE: 018501-000710US  
; CURRENT APPLICATION NUMBER: US/10/021.660  
; CURRENT FILING DATE: 2001-12-06  
; PRIOR APPLICATION NUMBER: US/09/784,356  
; PRIOR FILING DATE: 2001-02-14  
; PRIOR APPLICATION NUMBER: US 09/637,977  
; PRIOR FILING DATE: 2000-08-11  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 34  
; LENGTH: 3845  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-021-660-34

Query Match 80.0%; Score 15; DB 15; Length 3845;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGCCAGCCCCACAA 20  
|||  
Db 693 GCGCCAGCCCCACAA 678

## RESULT 30

US-10-116-802-162/c  
; Sequence 162, Application US/10116802  
; Publication No. US20030065157A1  
; GENERAL INFORMATION:  
; APPLICANT: Amy Lasek  
; TITLE OF INVENTION: GENES EXPRESSED IN LUNG CANCER  
; FILE REFERENCE: PA-0045 US  
; CURRENT APPLICATION NUMBER: US/10/116.802  
; CURRENT FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: 60/281,593  
; PRIOR FILING DATE: 2001-04-04  
; NUMBER OF SEQ ID NOS: 519  
; SOFTWARE: PERL Program  
; SEQ ID NO 162  
; LENGTH: 3914  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID NO: 3331519CB1  
US-10-116-802-162

Query Match 80.0%; Score 16; DB 13; Length 3914;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGCCAGCCCCACAA 20  
|||  
Db 754 GCGCCAGCCCCACAA 739

Search completed: June 20, 2004, 17:44:06  
Job time : 127.122 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 29.8571 Seconds

(without alignments)  
501.846 Million cell updates/sec

Title: US-10-624-714-12

Perfect score: 27

Sequence: 1 caccgcgcgttagatcgacactga 27

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	100.0	4403765	3	US-09-103-840A-2
2	27	100.0	4411529	3	US-09-103-840A-1
3	17.4	64.4	4403765	3	US-09-103-840A-2
4	17.4	64.4	4411529	3	US-09-103-840A-1
5	17	63.0	1701	2	US-08-484-938B-40
6	17	63.0	1701	2	US-08-484-938B-40
7	17	63.0	1701	2	US-08-484-938B-40
8	17	63.0	1701	2	US-08-484-938B-40
9	17	63.0	1701	3	US-08-484-938B-40
10	17	63.0	1701	3	US-08-484-938B-40
11	16.6	61.5	2262	4	US-09-149-223A-40
12	16.6	61.5	2262	4	US-09-149-223A-40
13	16.4	60.7	102	1	US-08-044-153D-37
14	16.4	60.7	235	1	US-07-743-518-21
15	16.4	60.7	427	1	US-08-014-153D-5
16	16.4	60.7	520	1	US-07-743-518-25
17	16.4	60.7	723	4	US-09-107-532A-1993
18	16.4	60.7	1317	4	US-09-107-532A-1993
19	16.4	60.7	1814	5	PCT-US96-00547-54
20	16.4	60.7	1892	5	PCT-US96-00547-54
21	16.4	60.7	2159	3	US-08-286-870A-7
22	16.4	60.7	9045	3	US-09-121-321-1
23	16.4	60.7	9045	3	US-08-933-803A-1
24	16.4	60.7	319608	4	US-09-539-333D-1
25	16.4	60.7	319608	4	US-09-539-333D-1
26	16.4	60.7	1830121	4	US-09-557-884-1
27	16.4	60.7	1830121	4	US-09-557-884-1

1605	60.0	16.2	28	US-09-107-532A-2895	Sequence 2895, Ap
405	59.3	16	29	US-09-252-991A-12862	Sequence 12862, A
551	59.3	16	30	US-09-489-039A-888	Sequence 888, Ap
866	59.3	16	31	US-08-484-938B-46	Sequence 46, Appl
866	59.3	16	32	US-08-484-158B-46	Sequence 46, Appl
866	59.3	16	33	US-08-484-596A-46	Sequence 46, Appl
866	59.3	16	34	US-08-480-150A-46	Sequence 46, Appl
866	59.3	16	35	US-08-458-731-46	Sequence 46, Appl
866	59.3	16	36	US-08-149-223A-46	Sequence 46, Appl
292	59.3	16	37	US-09-023-655-936	Sequence 936, Appl
2430	59.3	16	38	US-09-252-991A-12709	Sequence 12709, A
2607	59.3	16	39	US-09-252-991A-13249	Sequence 13249, A
2880	59.3	16	40	US-08-462-484-3	Sequence 3, Appl
2880	59.3	16	41	US-08-441-147-3	Sequence 3, Appl
2880	59.3	16	42	PCT-US95-07536-3	Sequence 3, Appl
3243	59.3	16	43	US-09-543-681A-1875	Sequence 1875, Ap
4021	59.3	16	44	US-09-453-702B-197	Sequence 197, Ap
4156	59.3	16	45	US-08-465-687A-1	Sequence 1, Appl
4156	59.3	16	46	US-09-030-970-1	Sequence 1, Appl
4156	59.3	16	47	US-09-520-210-1	Sequence 1, Appl
4156	59.3	16	48	PCT-US94-11843-1	Sequence 1, Appl
309	58.5	15.8	49	US-09-540-236-1768	Sequence 1768, Ap
591	58.5	15.8	50	US-09-328-352-2114	Sequence 2114, Ap
765	58.5	15.8	51	US-09-328-352-351	Sequence 391, Appl
795	58.5	15.8	52	US-09-543-681A-2954	Sequence 2954, Ap
1399	58.5	15.8	53	US-08-617-785-5	Sequence 5, Appl
1399	58.5	15.8	54	US-09-817-464-5	Sequence 5, Appl
3804	58.5	15.8	55	US-08-617-785-3	Sequence 3, Appl
3804	58.5	15.8	56	US-09-817-464-3	Sequence 3, Appl
89047	58.5	15.8	57	US-09-596-002-34	Sequence 34, Appl
1230025	58.5	15.8	58	US-09-198-452A-1	Sequence 1, Appl
28	57.8	15.6	59	US-08-859-998-150	Sequence 150, Ap
28	57.8	15.6	60	US-09-225-928-150	Sequence 150, Ap
28	57.8	15.6	61	US-09-225-201B-150	Sequence 150, Ap
270	57.8	15.6	62	US-09-313-294A-662	Sequence 662, Ap
313	57.8	15.6	63	US-09-313-294A-3867	Sequence 3867, Ap
360	57.8	15.6	64	US-08-256-077-3	Sequence 3, Appl
360	57.8	15.6	65	US-08-466-127-3	Sequence 3, Appl
684	57.8	15.6	66	US-09-489-039A-3021	Sequence 3021, Ap
735	57.8	15.6	67	US-09-252-991A-9675	Sequence 9675, Ap
756	57.8	15.6	68	US-07-717-332D-1	Sequence 1, Appl
939	57.8	15.6	69	US-09-252-991A-13966	Sequence 13966, A
982	57.8	15.6	70	US-08-256-077-1	Sequence 1, Appl
982	57.8	15.6	71	US-08-466-127-1	Sequence 1, Appl
1047	57.8	15.6	72	US-09-252-991A-13479	Sequence 13479, A
1512	57.8	15.6	73	US-09-543-681A-3340	Sequence 3340, Ap
1539	57.8	15.6	74	US-09-252-991A-9575	Sequence 9575, Ap
1797	57.8	15.6	75	US-09-252-991A-9716	Sequence 9716, Ap
3613	57.8	15.6	76	US-09-634-238-120	Sequence 120, Ap
7669	57.8	15.6	77	US-09-454-071-1	Sequence 1, Appl
80161	57.8	15.6	78	US-09-036-987A-1	Sequence 1, Appl
80161	57.8	15.6	79	US-09-370-700-1	Sequence 1, Appl
80161	57.8	15.6	80	US-09-603-207-1	Sequence 1, Appl
210	57.0	15.4	81	US-09-107-532A-747	Sequence 747, Ap
273	57.0	15.4	82	US-09-313-294A-1679	Sequence 1679, Ap
316	57.0	15.4	83	US-09-313-294A-5769	Sequence 5769, Ap
447	57.0	15.4	84	US-09-252-991A-14604	Sequence 14604, A
480	57.0	15.4	85	US-09-020-956-91	Sequence 91, Appl
480	57.0	15.4	86	US-09-030-607-91	Sequence 91, Appl
480	57.0	15.4	87	US-09-439-313-91	Sequence 91, Appl
480	57.0	15.4	88	US-09-352-616A-91	Sequence 91, Appl
480	57.0	15.4	89	US-09-232-149A-91	Sequence 91, Appl
480	57.0	15.4	90	US-09-159-812-91	Sequence 91, Appl
480	57.0	15.4	91	US-09-636-215-91	Sequence 91, Appl
480	57.0	15.4	92	US-09-683-168A-91	Sequence 91, Appl
480	57.0	15.4	93	US-09-115-453-91	Sequence 91, Appl
480	57.0	15.4	94	US-09-688-489-91	Sequence 91, Appl
601	57.0	15.4	95	US-09-030-607-192	Sequence 192, App
601	57.0	15.4	96	US-09-439-313-192	Sequence 192, App
601	57.0	15.4	97	US-09-352-616A-192	Sequence 192, App
601	57.0	15.4	98	US-09-232-149A-192	Sequence 192, App
601	57.0	15.4	99	US-09-153-812-192	Sequence 192, App
601	57.0	15.4	100	US-09-636-215-192	Sequence 192, App

C 101	15.4	57.0	601	4	US-09-685-166A-192	Sequence 192, App	C 174	15	55.6	2040	4	US-09-639-206A-5	Sequence 5, Appli
C 102	15.4	57.0	601	4	US-09-115-453-192	Sequence 192, App	C 175	15	55.6	2040	4	US-09-874-923-5	Sequence 5, Appli
C 103	15.4	57.0	601	4	US-09-688-489-192	Sequence 192, App	C 176	15	55.6	2163	6	5281520-1	Patent No. 5281520
C 104	15.4	57.0	609	4	US-09-252-991A-14388	Sequence 14388, A	C 177	15	55.6	2163	6	5281520-2	Patent No. 5281520
C 105	15.4	57.0	624	4	US-09-221-017B-43	Sequence 43, Appl	C 178	15	55.6	2291	6	5281520-3	Sequence 14064, A
C 106	15.4	57.0	633	4	US-09-252-991A-14315	Sequence 14315, A	C 179	15	55.6	2298	4	US-09-252-991A-14064	Sequence 5, Appli
C 107	15.4	57.0	777	4	US-09-252-991A-9775	Sequence 9775, Ap	C 180	15	55.6	2565	1	US-08-619-554-5	Sequence 1, Appli
C 108	15.4	57.0	825	4	US-09-107-532A-1310	Sequence 1310, Ap	C 181	15	55.6	4322	2	US-08-537-343-1	Sequence 29, Appl
C 109	15.4	57.0	924	4	US-09-252-991A-15149	Sequence 15149, A	C 182	15	55.6	5138	4	US-09-526-193A-29	Sequence 29, Appl
C 110	15.4	57.0	1134	4	US-09-711-164-279	Sequence 279, App	C 183	15	55.6	7393	4	US-09-620-312D-372	Sequence 372, App
C 111	15.4	57.0	1206	4	US-09-252-991A-2124	Sequence 2124, Ap	C 184	15	55.6	7860	4	US-09-526-193A-29	Sequence 2, Appli
C 112	15.4	57.0	1306	4	US-09-540-236-1386	Sequence 1386, Ap	C 185	15	55.6	35881	4	US-08-311-731A-127	Sequence 127, App
C 113	15.4	57.0	1306	4	US-09-252-991A-14284	Sequence 14284, A	C 186	15	55.6	536165	4	US-09-214-808-1	Sequence 1, Appli
C 114	15.4	57.0	1380	4	US-08-426-630-15	Sequence 15, Appl	C 187	14.8	54.8	36	1	US-08-182-530-12	Sequence 12, Appl
C 115	15.4	57.0	1629	4	US-08-396-452-4	Sequence 2042, Ap	C 188	14.8	54.8	36	1	US-08-050-058B-12	Sequence 12, Appl
C 116	15.4	57.0	1705	1	US-09-169-119-4	Sequence 4, Appli	C 189	14.8	54.8	36	1	US-08-463-587A-11	Sequence 11, Appl
C 117	15.4	57.0	1705	3	US-09-252-991A-8888	Sequence 4, Appli	C 190	14.8	54.8	36	2	US-08-441-871-17	Sequence 17, Appl
C 118	15.4	57.0	1857	4	US-09-252-991A-9114	Sequence 8888, Ap	C 191	14.8	54.8	36	3	US-08-923-854-11	Sequence 11, Appl
C 119	15.4	57.0	1908	4	US-09-252-991A-9114	Sequence 1914, Ap	C 192	14.8	54.8	36	5	US-09-543-681A-901	Sequence 12, Appl
C 120	15.4	57.0	3725	1	US-08-155-331-12	Sequence 12, Appl	C 193	14.8	54.8	186	4	US-09-313-294A-6266	Sequence 6266, Ap
C 121	15.4	57.0	3725	1	US-08-424-022-12	Sequence 12, Appl	C 194	14.8	54.8	300	4	US-09-252-991A-5752	Sequence 5752, Ap
C 122	15.4	57.0	3725	2	US-08-424-017B-12	Sequence 12, Appl	C 195	14.8	54.8	489	4	US-09-621-976-10097	Sequence 10097, A
C 123	15.4	57.0	3725	5	US-09-252-991A-9183	Sequence 12, Appl	C 196	14.8	54.8	502	4	US-09-669-751-17	Sequence 17, Appl
C 124	15.4	57.0	6573	4	US-08-426-630-2	Sequence 9183, Ap	C 197	14.8	54.8	581	4	US-09-364-206-10	Sequence 10, Appl
C 125	15.4	57.0	8753	4	US-08-426-630-2	Sequence 2, Appli	C 198	14.8	54.8	586	4	US-09-669-751-19	Sequence 19, Appl
C 126	15.4	57.0	99629	4	US-09-596-003-37	Sequence 37, Appl	C 199	14.8	54.8	617	4	US-09-669-751-19	Sequence 19, Appl
C 127	15.4	57.0	1230025	4	US-09-198-452A-1	Sequence 1, Appli	C 200	14.8	54.8	632	4	US-09-669-751-19	Sequence 19, Appl
C 128	15.2	56.3	486	4	US-09-252-991A-7451	Sequence 7451, Ap	C 201	14.8	54.8	635	3	US-09-998-416-110	Sequence 110, Appl
C 129	15.2	56.3	524	4	US-09-621-976-14012	Sequence 14012, A	C 202	14.8	54.8	678	4	US-09-489-039A-2981	Sequence 2981, Ap
C 130	15.2	56.3	1242	4	US-09-489-039A-5364	Sequence 5364, Ap	C 203	14.8	54.8	680	4	US-09-556-877-267	Sequence 267, App
C 131	15.2	56.3	1464	3	US-09-188-930-255	Sequence 255, App	C 204	14.8	54.8	680	4	US-09-520-413C-267	Sequence 267, App
C 132	15.2	56.3	1464	4	US-09-312-283C-255	Sequence 255, App	C 205	14.8	54.8	680	4	US-09-598-413-267	Sequence 267, App
C 133	15.2	56.3	1491	2	US-08-310-912A-105	Sequence 105, App	C 206	14.8	54.8	711	3	US-08-935-263-1	Sequence 1, Appli
C 134	15.2	56.3	1491	2	US-08-841-089-105	Sequence 105, App	C 207	14.8	54.8	711	4	US-09-594-185-1	Sequence 10, Appl
C 135	15.2	56.3	1491	3	US-09-301-085-105	Sequence 105, App	C 208	14.8	54.8	810	4	US-08-778-717-10	Sequence 10, Appl
C 136	15.2	56.3	1491	5	US-09-301-085-105	Sequence 105, App	C 209	14.8	54.8	810	4	US-09-489-039A-6447	Sequence 6447, Ap
C 137	15.2	56.3	1491	5	US-09-301-085-105	Sequence 105, App	C 210	14.8	54.8	810	4	US-09-403-768-3	Sequence 3, Appli
C 138	15.2	56.3	1633	3	US-09-188-930-73	Sequence 73, Appl	C 211	14.8	54.8	1038	4	US-09-252-991A-16298	Sequence 16298, A
C 139	15.2	56.3	1635	4	US-09-312-283C-73	Sequence 73, Appl	C 212	14.8	54.8	1050	4	US-09-543-681A-4077	Sequence 4077, Ap
C 140	15.2	56.3	2030	2	US-08-923-536A-1	Sequence 1, Appli	C 213	14.8	54.8	1173	4	US-09-252-991A-15979	Sequence 15979, A
C 141	15.2	56.3	3729	2	US-08-680-326-26	Sequence 26, Appl	C 214	14.8	54.8	1401	4	US-09-134-001C-595	Sequence 595, App
C 142	15.2	56.3	3729	4	US-09-904-065-11	Sequence 11, Appl	C 215	14.8	54.8	1566	4	US-09-489-039A-6221	Sequence 706, App
C 143	15.2	56.3	4657	3	US-09-254-325-1	Sequence 1, Appli	C 216	14.8	54.8	1731	4	US-09-328-352-706	Sequence 688, App
C 144	15.2	56.3	5643	3	US-09-079-415-5	Sequence 5, Appli	C 217	14.8	54.8	1731	4	US-09-489-039A-688	Sequence 83, Appl
C 145	15.2	56.3	5643	3	US-08-750-458A-1	Sequence 1, Appli	C 218	14.8	54.8	1837	4	US-09-220-132-83	Sequence 3, Appli
C 146	15.2	56.3	5903	4	US-09-453-702B-195	Sequence 195, App	C 219	14.8	54.8	1837	4	US-09-393-839-3	Sequence 3, Appli
C 147	15.2	56.3	6306	1	US-08-466-390-3	Sequence 3, Appli	C 220	14.8	54.8	2152	4	US-09-449-632-3	Sequence 2033, Ap
C 148	15.2	56.3	6306	1	US-08-470-950-3	Sequence 3, Appli	C 221	14.8	54.8	2298	4	US-09-489-039A-2033	Sequence 8696, Ap
C 149	15.2	56.3	6306	1	US-08-467-781-3	Sequence 3, Appli	C 222	14.8	54.8	2313	4	US-09-252-991A-8696	Sequence 5660, Ap
C 150	15.2	56.3	6306	1	US-08-195-487-3	Sequence 3, Appli	C 223	14.8	54.8	2367	4	US-09-489-039A-5560	Sequence 11, Appl
C 151	15.2	56.3	6306	1	US-08-483-924-3	Sequence 3, Appli	C 224	14.8	54.8	2549	1	US-07-653-091A-11	Sequence 11, Appl
C 152	15.2	56.3	6306	5	US-08-483-924-3	Sequence 3, Appli	C 225	14.8	54.8	2549	1	US-08-481-598-11	Sequence 11, Appl
C 153	15.2	56.3	36138	5	US-08-311-731A-136	Sequence 136, App	C 226	14.8	54.8	2549	2	US-08-481-598-11	Sequence 11, Appl
C 154	15	55.6	822	4	US-09-252-991A-1046	Sequence 1046, Ap	C 227	14.8	54.8	2549	2	US-08-483-353-11	Sequence 11, Appl
C 155	15	55.6	861	4	US-09-252-991A-13574	Sequence 13574, A	C 228	14.8	54.8	2549	2	US-08-473-702-11	Sequence 11, Appl
C 156	15	55.6	861	4	US-09-252-991A-13772	Sequence 13772, A	C 229	14.8	54.8	2639	6	US-07-952-817-8	Sequence 8, Appli
C 157	15	55.6	954	4	US-09-489-039A-3455	Sequence 3455, Ap	C 230	14.8	54.8	2639	6	US-09-252-991A-9082	Sequence 9082, Ap
C 158	15	55.6	1071	4	US-09-252-991A-14206	Sequence 14206, A	C 231	14.8	54.8	2706	4	US-09-107-532A-183	Sequence 183, App
C 159	15	55.6	1182	4	US-09-165-827C-13	Sequence 13, Appli	C 232	14.8	54.8	2709	4	US-08-091-569-1	Sequence 1, Appli
C 160	15	55.6	1185	4	US-09-165-827C-13	Sequence 13, Appli	C 233	14.8	54.8	3695	1	US-08-403-676-1	Sequence 1, Appli
C 161	15	55.6	1212	6	US-09-252-991A-13872	Sequence 13872, A	C 234	14.8	54.8	3695	2	US-08-822-238-1	Sequence 1, Appli
C 162	15	55.6	1305	4	US-09-252-991A-13872	Sequence 13872, A	C 235	14.8	54.8	3695	2	US-08-822-238-1	Sequence 1, Appli
C 163	15	55.6	1332	1	US-08-133-347-1	Sequence 1, Appli	C 236	14.8	54.8	4084	3	US-09-103-875-4	Sequence 4, Appli
C 164	15	55.6	1635	1	US-08-133-347-1	Sequence 1, Appli	C 237	14.8	54.8	4460	3	US-09-103-875-4	Sequence 1, Appli
C 165	15	55.6	1743	6	US-09-252-991A-10066	Sequence 10066, A	C 238	14.8	54.8	4946	3	US-09-300-996A-1	Sequence 7, Appli
C 166	15	55.6	1756	6	US-09-252-991A-10066	Sequence 10066, A	C 239	14.8	54.8	5570	3	US-09-066-046-7	Sequence 4, Appli
C 167	15	55.6	1998	6	US-09-252-991A-10066	Sequence 10066, A	C 240	14.8	54.8	6085	3	US-09-029-603-4	Sequence 4, Appli
C 168	15	55.6	2040	2	US-08-533-669A-5	Sequence 5, Appli	C 241	14.8	54.8	6133	4	US-09-453-702B-15	Sequence 15, Appl
C 169	15	55.6	2040	4	US-09-183-861-5	Sequence 5, Appli	C 242	14.8	54.8	8509	4	US-09-826-205-1	Sequence 1, Appli
C 170	15	55.6	2040	4	US-09-022-765-5	Sequence 5, Appli	C 243	14.8	54.8	34094	4	US-09-292-034-1	Sequence 1, Appli
C 171	15	55.6	2040	4	US-09-551-974A-5	Sequence 5, Appli	C 244	14.8	54.8	35081	2	US-08-752-760A-1	Sequence 134, App
C 172	15	55.6	2040	4	US-09-551-974A-5	Sequence 5, Appli	C 245	14.8	54.8	36241	4	US-08-311-731A-134	Sequence 1, Appli
C 173	15	55.6	2040	4	US-09-565-501A-5	Sequence 5, Appli	C 246	14.8	54.8	38653	4	US-09-922-445-1	Sequence 1, Appli



c 247 14.6 54.1 219 4 US-09-489-039A-6610  
248 14.6 54.1 255 4 US-09-489-039A-6822  
249 14.6 54.1 306 4 US-09-313-294A-7024  
c 250 14.6 54.1 378 1 US-08-202-047-12  
251 14.6 54.1 378 1 US-08-202-047-14  
c 252 14.6 54.1 378 1 US-08-202-047-16  
253 14.6 54.1 378 1 US-08-202-047-18  
c 254 14.6 54.1 378 3 US-08-964-690-12  
255 14.6 54.1 378 3 US-08-964-690-14  
c 256 14.6 54.1 378 3 US-08-964-690-16  
257 14.6 54.1 378 3 US-08-964-690-18  
c 258 14.6 54.1 402 4 US-09-408-020-35  
259 14.6 54.1 402 4 US-09-408-020-69  
c 260 14.6 54.1 585 4 US-09-252-931A-7116  
261 14.6 54.1 753 4 US-09-252-931A-14200  
c 262 14.6 54.1 942 4 US-09-252-931A-7060  
263 14.6 54.1 1164 4 US-09-252-931A-4972  
264 14.6 54.1 1218 4 US-09-489-039A-2871  
265 14.6 54.1 1248 4 US-09-489-039A-3687  
266 14.6 54.1 1332 4 US-09-252-931A-4992  
267 14.6 54.1 1407 4 US-09-252-931A-13987  
268 14.6 54.1 1410 4 US-09-252-931A-14025  
269 14.6 54.1 1413 4 US-09-489-039A-7120  
270 14.6 54.1 1419 4 US-09-252-931A-6988  
c 271 14.6 54.1 1515 4 US-09-543-681A-2349  
272 14.6 54.1 1536 4 US-09-489-039A-3862  
c 273 14.6 54.1 1743 4 US-09-489-039A-7034  
c 274 14.6 54.1 1908 4 US-09-252-931A-4943  
c 275 14.6 54.1 2373 4 US-09-252-931A-16481  
c 276 14.6 54.1 3424 4 US-09-553-690-3  
277 14.6 54.1 5410 4 US-09-221-017B-70  
278 14.6 54.1 6250 3 US-08-729-214-23  
279 14.6 54.1 6250 3 US-09-028-934-23  
280 14.6 54.1 7001 1 US-08-258-261B-1  
281 14.6 54.1 7001 1 US-08-456-837-1  
282 14.6 54.1 7001 1 US-08-457-342-1  
283 14.6 54.1 7001 1 US-08-457-646A-1  
284 14.6 54.1 7001 1 US-08-458-076A-1  
285 14.6 54.1 7001 1 US-08-457-335A-1  
286 14.6 54.1 7001 1 US-08-761-258-6  
287 14.6 54.1 7001 1 US-08-723-214-1  
288 14.6 54.1 7001 2 US-08-977-306-6  
289 14.6 54.1 7001 3 US-09-028-934-1  
c 290 14.6 54.1 7278 3 US-09-091-219-1  
c 291 14.6 54.1 7278 4 US-09-660-541-1  
c 292 14.6 54.1 7778 4 US-08-956-171E-48  
293 14.6 54.1 9495 1 US-08-271-829-1  
c 294 14.6 54.1 9495 5 PCT-US93-01544-1  
c 295 14.6 54.1 12785 4 US-09-553-690-1  
296 14.6 54.1 32998 4 US-09-408-020-1  
c 297 14.6 54.1 37030 4 US-08-311-731A-25  
298 14.6 54.1 42432 4 US-09-408-020-2  
299 14.4 53.3 284 4 US-09-313-294A-4622  
300 14.4 53.3 348 4 US-09-328-352-3669

## ALIGNMENTS

US-09-103-840A-2/c  
RESULT 1  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 100.0%; Score 27; DB 3; Length 4403765;  
Best Local Similarity 100.0%; Pred. No. 0.0015;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CACCCGGCGTTTAGGATCGACACCTGA 27  
|||||  
Db 2600107 CACCCGGCGTTTAGGATCGACACCTGA 2600081

## RESULT 2

US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37RV  
US-09-103-840A-1

Query Match 100.0%; Score 27; DB 3; Length 4411529;  
Best Local Similarity 100.0%; Pred. No. 0.0015;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CACCCGGCGTTTAGGATCGACACCTGA 27  
|||||  
Db 2604297 CACCCGGCGTTTAGGATCGACACCTGA 2604271

## RESULT 3

US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis

FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 64.4%; Score 17.4; DB 3; Length 4403765;  
Best Local Similarity 77.8%; Pred. No. 49;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27  
DB 4043594 CCCCCGCGCCAGGATCGTCACCTTA 4043620

RESULT 4  
US-09-103-840A-1  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT FILING DATE: 1998-06-24  
CURRENT APPLICATION NUMBER: US/09/103,840A  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37RV  
US-09-103-840A-1

Query Match 64.4%; Score 17.4; DB 3; Length 4411529;  
Best Local Similarity 77.8%; Pred. No. 49;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27  
DB 4051526 CCCCCGCGCCAGGATCGTCACCTTA 4051552

RESULT 5  
US-08-484-993B-40  
Sequence 40, Application US/08484993B  
Patent No. 5837497  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Jeffrey D.  
APPLICANT: Hsu, Kuang T.  
APPLICANT: Podolski, Joseph S.  
TITLE OF INVENTION: Materials and Methods for Immunocontraception  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,993B  
FILING DATE: 09-NOV-1993  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/012,990  
FILING DATE: 29-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,341  
FILING DATE: 09-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 31745  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6653  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1701 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1698  
US-08-484-993B-40

Query Match 63.0%; Score 17; DB 2; Length 1701;  
Best Local Similarity 80.0%; Pred. No. 43;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCT 25  
DB 1318 CACGCGCTTCAGCATCTTCACCT 1342

RESULT 6  
US-08-484-158B-40  
Sequence 40, Application US/08484158B  
Patent No. 5976545  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Jeffrey D.  
APPLICANT: Hsu, Kuang T.  
APPLICANT: Podolski, Joseph S.  
TITLE OF INVENTION: Pharmaceutical Compositions for Immunocontraception  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,158B  
FILING DATE: 07-JUNE-95  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/149,223  
FILING DATE: 09-NOV-93  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/012,990  
FILING DATE: 29-JAN-93  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,341  
FILING DATE: 09-NOV-92  
ATTORNEY/AGENT INFORMATION:

NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 32794  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6653  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1701 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1698  
US-08-484-158B-40

Query Match 63.0%; Score 17; DB 2; Length 1701;  
Best Local Similarity 80.0%; Pred. No. 43;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CACCGCGCTTTAGATCGACACCT 25  
Db 1318 CACGAGCGCTTCAGCATCTTCACCT 1342

RESULT 7  
US-08-484-596A-40  
Sequence 40, Application US/08484596A  
Patent No. 5981228  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Jeffrey D.  
APPLICANT: Hsu, Kuang T.  
APPLICANT: Podolski, Joseph S.  
TITLE OF INVENTION: Materials and Methods for Immunocontraception  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,596A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION NUMBER: 08/149,223  
FILING DATE: 09-NOV-1992  
APPLICATION NUMBER: 08/149,223  
FILING DATE: 09-NOV-1992  
APPLICATION NUMBER: 08/012,990  
FILING DATE: 29-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,341  
FILING DATE: 09-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 31745  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6653  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1701 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1698  
US-08-484-150A-40

Query Match 63.0%; Score 17; DB 2; Length 1701;  
Best Local Similarity 80.0%; Pred. No. 43;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1698  
US-08-484-596A-40

Query Match 63.0%; Score 17; DB 2; Length 1701;  
Best Local Similarity 80.0%; Pred. No. 43;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CACCGCGCTTTAGATCGACACCT 25  
Db 1318 CACGAGCGCTTCAGCATCTTCACCT 1342

RESULT 8  
US-08-480-150A-40  
Sequence 40, Application US/08480150A  
Patent No. 5989550  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Jeffrey D.  
APPLICANT: Hsu, Kuang T.  
APPLICANT: Podolski, Joseph S.  
TITLE OF INVENTION: Materials and Methods for Immunocontraception  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,150A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/149,223  
FILING DATE: 09-NOV-1993  
APPLICATION NUMBER: 08/012,990  
FILING DATE: 29-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,341  
FILING DATE: 09-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 31745  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6653  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1701 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1698  
US-08-480-150A-40

Query Match 63.0%; Score 17; DB 2; Length 1701;  
Best Local Similarity 80.0%; Pred. No. 43;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

us-10-624-714-12.rni

Mon Jun 21 09:02:06 2004

```

;
; APPLICANT: Hsu, Kuang T.
; APPLICANT: Podolski, Joseph S.
; TITLE OF INVENTION: Materials and Methods for Immunocontraception
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/149,223A
; FILING DATE: 09-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/012,990
; FILING DATE: 29-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,341
; FILING DATE: 09-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 31745
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6653
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1698
; US-08-149-223A-40
;
; Query Match 63.0%; Score 17; DB 3; Length 1701;
; Best Local Similarity 80.0%; Pred. No. 43;
; Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCT 25
Db 1318 CACACGCGCTTCAGCATCTTCACCT 1342

RESULT 11
US-09-252-991A-1416/c
; Sequence 1416; Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; ADDRESSEE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1416
; LENGTH: 2262

;
; APPLICANT: Harris Ph.D., Jeffrey D.
; APPLICANT: Hsu, Kuang T.
; APPLICANT: Podolski, Joseph S.
; TITLE OF INVENTION: Materials and Methods for Immunocontraception
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,731
; FILING DATE: 09-NOV-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/012,990
; FILING DATE: 29-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,341
; FILING DATE: 09-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 31745
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6653
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1698
; US-08-458-731-40
;
; Query Match 63.0%; Score 17; DB 3; Length 1701;
; Best Local Similarity 80.0%; Pred. No. 43;
; Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCT 25
Db 1318 CACACGCGCTTCAGCATCTTCACCT 1342

RESULT 10
US-08-149-223A-40
; Sequence 40; Application US/08149223A
; Patent No. 6027727
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Jeffrey D.

```

FILING DATE: 13-JUN-1989  
APPLICATION NUMBER: US 06/948,270  
FILING DATE: 31-DEC-1986  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, Allan A.  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: G4C1P3  
TELEPHONE: (415) 369-3500  
TELEFAX: (415) 369-3500  
TELECOMMUNICATION INFORMATION:  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 102 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: 3A3B(I), Fig. 8  
SEQUENCE DESCRIPTION: SEQ ID NO: 37:  
US-08-014-153D-37  
Query Match 60.7%; Score 16.4; DB 1; Length 102;  
Best Local Similarity 76.9%; Pred. NO. 63;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
2 ACCCGCGCTTAGGATCGACACCTGA 27  
|||||  
60 ACCCGCGCTTAGGATCGACACCTGA 85  
RESULT 14  
US-07-743-518-21  
Sequence 21, Application US/07743518  
Patent No. 5397696  
GENERAL INFORMATION:  
APPLICANT: YAMAGIHARA, RICHARD  
APPLICANT: NERURKAR, VIVEK R.  
APPLICANT: JENKINS, CAROL  
APPLICANT: MILLER, MARK  
APPLICANT: GARUTO, RALPH M.  
TITLE OF INVENTION: PAPUA NEW GUINEA HUMAN T-LYMPHOTROPIC  
TITLE OF INVENTION: VIRUS  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
STREET: 1615 L STREET, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20036-5601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DCS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07743518  
FILING DATE: 19910812  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: SCOTT, WATSON I.  
REGISTRATION NUMBER: 26,581  
REFERENCE/DOCKET NUMBER: WTS/5683/84699/SAP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861-3000  
TELEFAX: 202-822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 235 base pairs  
TYPE: NUCLEIC ACID

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 32..205  
US-07-743-518-21

Query Match 60.7%; Score 16.4; DB 1; Length 235;  
Best Local Similarity 76.9%; Pred. No. 68;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACTGA 27  
DB 157 ACCCGCGCTTTAGGATCGAGTCTCTGA 182

# RESULT 15

US-08-014-153D-5  
Sequence 5, Application US/08014153D  
GENERAL INFORMATION:  
APPLICANT: Hadlock, Kenneth G.  
Goh, Chin-Joo

TITLE OF INVENTION: Method and Assay for HTLV  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genelabs Technologies, Inc.  
STREET: 505 Pembiscot Drive  
CITY: Redwood City  
STATE: CA  
COUNTRY: USA  
ZIP: 94063

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/014,153D  
FILING DATE: 05-Feb-1993  
CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/653,051  
FILING DATE: 08-FEB-1991  
APPLICATION NUMBER: US 07/366,313  
FILING DATE: 13-JUN-1989  
APPLICATION NUMBER: US 06/948,270  
FILING DATE: 31-DEC-1986

## ATTORNEY/AGENT INFORMATION:

NAME: Brookes, Allan A.  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: G4C1P3  
TELEPHONE: (415) 369-9500  
TELEFAX: (415) 368-0709

## INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
LENGTH: 427 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: p21E(1)CS, Fig. 2A  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Query Match 60.7%; Score 16.4; DB 1; Length 427;  
Best Local Similarity 76.9%; Pred. No. 73;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACTGA 27  
DB 354 ACCCGCGCTTTAGGATCGAGTCTCTGA 379

# RESULT 16

US-07-743-518-25  
Sequence 25, Application US/07743518  
Patent No. 5397696  
GENERAL INFORMATION:

APPLICANT: YANAGIHARA, RICHARD  
APPLICANT: NERURKAR, VIVEK R.  
APPLICANT: JENKINS, CAROL  
APPLICANT: MILLER, MARK  
APPLICANT: GARRUTO, RALPH M.  
TITLE OF INVENTION: PAPUA NEW GUINEA HUMAN T-LYMPHOTROPIC  
TITLE OF INVENTION: VIRUS  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
STREET: 1615 L STREET, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20036-5601

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/743,518  
FILING DATE: 19910812  
CLASSIFICATION: 424

## ATTORNEY/AGENT INFORMATION:

NAME: SCOTT, WATSON T.  
REGISTRATION NUMBER: 26,581  
REFERENCE/DOCKET NUMBER: WTS/5683/84699/SAP  
TELEPHONE: 202-861-3000  
TELEFAX: 202-822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 25:

## SEQUENCE CHARACTERISTICS:

LENGTH: 520 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA

## US-07-743-518-25

Query Match 60.7%; Score 16.4; DB 1; Length 520;  
Best Local Similarity 76.9%; Pred. No. 74;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACTGA 27  
DB 382 ACCCGCGCTTTAGGATCGAGTCTCTGA 407

# RESULT 17

US-09-107-532A-1993  
Sequence 1993, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham

STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICANT: Virogenetics Corporation  
TITLE OF INVENTION: RECOMBINANT POXVIRUS-HTLV, COMPOSITIONS  
FILE REFERENCE: 2709.1002-001  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US/09/543,681A  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 613  
LENGTH: 1317  
TYPE: DNA  
ORGANISM: Proteus mirabilis  
US-09-543-681A-613

Query Match 60.7%; Score 16.4; DB 4; Length 1317;  
Best Local Similarity 76.9%; Pred. No. 82;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 10 TTTAGGATCGACACTGA 27  
Db 594 TTTAGGATCGACACTGA 611

RESULT 18  
US-09-543-681A-613  
; Sequence 613, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 613  
; LENGTH: 1317  
; TYPE: DNA  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-613

Query Match 60.7%; Score 16.4; DB 4; Length 1317;  
Best Local Similarity 76.9%; Pred. No. 82;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCGCGCTTTAGGATCGACACTG 26  
Db 1158 CATCCCGCATTTAGGATCGTCAATG 1183

RESULT 19  
PCT-US96-00547-54  
; Sequence 54, Application PC/TUS9600547  
; GENERAL INFORMATION:  
; APPLICANT: Virogenetics Corporation  
; TITLE OF INVENTION: RECOMBINANT POXVIRUS-HTLV, COMPOSITIONS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/107,532A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 613  
; LENGTH: 1317  
; TYPE: DNA  
; ORGANISM: Proteus mirabilis  
PCT-US96-00547-54

Query Match 60.7%; Score 16.4; DB 5; Length 1814;  
Best Local Similarity 76.9%; Pred. No. 85;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 ACCCGCGCTTTAGGATCGACACTGA 27  
Db 1455 ACCCGCGCTTTAGGATCGAGTCTGTA 1480

RESULT 20  
PCT-US96-00547-41  
; Sequence 41, Application PC/TUS9600547  
; GENERAL INFORMATION:  
; APPLICANT: Virogenetics Corporation  
; TITLE OF INVENTION: RECOMBINANT POXVIRUS-HTLV, COMPOSITIONS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 613  
; LENGTH: 1317  
; TYPE: DNA  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-613

Query Match 60.7%; Score 16.4; DB 4; Length 1317;  
Best Local Similarity 76.9%; Pred. No. 82;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00547  
FILING DATE: 12-JAN-1996  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/372,664  
FILING DATE: 13-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2621  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1892 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US96-00547-41

Query Match 60.7%; Score 16.4; DB 5; Length 1892;  
Best Local Similarity 76.9%; Pred. No. 85;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACCTGA 27  
DB 1508 ACCCGCGCTTTAGGATCGACCTGA 1533

## RESULT 21

US-08-286-870A-7  
Sequence 7, Application US/08286870A  
Patent No. 6063605  
GENERAL INFORMATION:  
APPLICANT: ELY, S  
APPLICANT: TAILOR, RH  
APPLICANT: TIPPETT, JM  
APPLICANT: BLENK, RG  
TITLE OF INVENTION: BACTERIAL GENES  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DABY & CUSHMAN  
ADDRESSEE: Intellectual Property Group of  
ADDRESSEE: PILLSBURY, MADISON & SUTRO LLP  
STREET: 1100 New York Avenue, N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/286,870A  
FILING DATE: 05-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/520228  
FILING DATE: 09-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 8910624.9  
FILING DATE: 09-MAY-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: PAUL N. KOKULIS

REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 70608/220720  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2159 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2159  
US-08-286-870A-7

Query Match 60.7%; Score 16.4; DB 3; Length 2159;  
Best Local Similarity 76.9%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACCTG 26  
DB 916 CACCGCGCTTTAGGATCGACCTG 941

## RESULT 22

US-09-121-321-1  
Sequence 1, Application US/09121321  
Patent No. 6090783  
GENERAL INFORMATION:  
APPLICANT: Saiga, Akihiko  
APPLICANT: Orita, Satoshi  
APPLICANT: Igarashi, Hisanaga  
APPLICANT: Okumura, Kouichi  
APPLICANT: Sakaguchi, Gaku  
TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/121,321  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/933,803  
FILING DATE: 19-SEP-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley, James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: SHGN-12CIP  
TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9045 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:



ORGANISM: human  
FEATURE:  
NAME/KEY: LTR  
LOCATION: 1..757  
FEATURE: polyA signal  
NAME/KEY: polyA signal  
LOCATION: 8584..8589  
FEATURE:  
NAME/KEY: LTR  
LOCATION: 8278..9032  
US-09-121-321-1

Query Match 60.7%; Score 16.4; DB 3; Length 9045;  
Best Local Similarity 76.9%; Pred. No. 1e+02;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTGA 27  
DB 6427 ACCCGCGCTTTAGGATCGACACCTGA 6452

## RESULT 23

US-08-933-803A-1  
Sequence 1, Application US/08933803A  
Patent No. 6218522

GENERAL INFORMATION:  
APPLICANT: Saiga, Akihiko  
APPLICANT: Orita, Satoshi  
APPLICANT: Igarashi, Hisanaga  
APPLICANT: Okumura, Kouichi  
APPLICANT: Sakaguchi, Gaku  
TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF  
TITLE OF INVENTION: GENE EXPRESSION AND NOVEL PROTEIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESS: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/933,803A  
FILING DATE: 19-SEP-1997  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley, James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: SHGN-12CIP  
TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9045 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: human  
FEATURE:  
NAME/KEY: LTR  
LOCATION: 1..757  
FEATURE:  
NAME/KEY: polyA signal  
LOCATION: 8584..8589  
FEATURE:

NAME/KEY: LTR  
LOCATION: 8278..9032  
US-08-933-803A-1

Query Match 60.7%; Score 16.4; DB 3; Length 9045;  
Best Local Similarity 76.9%; Pred. No. 1e+02;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTGA 27  
DB 6427 ACCCGCGCTTTAGGATCGACACCTGA 6452

## RESULT 24

US-09-539-333D-1/c  
Sequence 1, Application US/09539333D  
Patent No. 6476208

GENERAL INFORMATION:  
APPLICANT: Cohen, Daniel  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Chumakov, Ilya  
APPLICANT: Bougueleret, Lydie  
APPLICANT: Bihain, Bernard  
APPLICANT: Essioux, Laurent  
TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS  
FILE REFERENCE: GENSET-047AUS  
CURRENT APPLICATION NUMBER: US/09/539,333D  
CURRENT FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: US 60/126,903  
PRIOR FILING DATE: 1999-03-30  
PRIOR APPLICATION NUMBER: US 60/131,971  
PRIOR FILING DATE: 1999-04-30  
PRIOR APPLICATION NUMBER: US 60/132,065  
PRIOR FILING DATE: 1999-04-30  
PRIOR APPLICATION NUMBER: US 60/143,928  
PRIOR FILING DATE: 1999-07-14  
PRIOR APPLICATION NUMBER: US 60/145,915  
PRIOR FILING DATE: 1999-07-27  
PRIOR APPLICATION NUMBER: US 60/146,453  
PRIOR FILING DATE: 1999-07-29  
PRIOR APPLICATION NUMBER: US 60/146,452  
PRIOR FILING DATE: 1999-07-29  
PRIOR APPLICATION NUMBER: US 60/162,288  
PRIOR FILING DATE: 1999-10-28  
PRIOR APPLICATION NUMBER: US 09/416,384  
PRIOR FILING DATE: 1999-10-12  
NUMBER OF SEQ ID NOS: 231  
SOFTWARE: Patent.pm  
SEQ ID NO 1

LENGTH: 319608  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 31..1107  
OTHER INFORMATION: 5'regulatory region g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1108..1289  
OTHER INFORMATION: exon A g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 14877..14920  
OTHER INFORMATION: exon B g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 18778..18862  
OTHER INFORMATION: exon Bbis g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 25593..25740  
OTHER INFORMATION: exon C g35018 gene  
FEATURE:

NAME/KEY: exon  
LOCATION: 29388..29502  
OTHER INFORMATION: exon D g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 29967..30282  
OTHER INFORMATION: exon E g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 64666..64812  
OTHER INFORMATION: exon F g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 65505..65853  
OTHER INFORMATION: exon G g35018 gene  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 65854..67854  
OTHER INFORMATION: 3'regulatory region g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 94124..94964  
OTHER INFORMATION: exon g35017  
FEATURE:  
NAME/KEY: exon  
LOCATION: 201188..201234  
OTHER INFORMATION: exon S g35030 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 214676..214793  
OTHER INFORMATION: exon T g35030 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 215702..215746  
OTHER INFORMATION: exon U g35030 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 216836..216915  
OTHER INFORMATION: exon V g35030 gene  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 213818..215818  
OTHER INFORMATION: 3'regulatory region g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 215819..215941  
OTHER INFORMATION: exon R complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 215819..215975  
OTHER INFORMATION: exon Rbis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 216661..216952  
OTHER INFORMATION: exon Qbis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 216661..217061  
OTHER INFORMATION: exon Q complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 217027..217061  
OTHER INFORMATION: exon Q1 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 229647..229742  
OTHER INFORMATION: exon X complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 230408..230721  
OTHER INFORMATION: exon P complement g34872 gene  
FEATURE:  
NAME/KEY: exon

LOCATION: 231272..231412  
OTHER INFORMATION: exon Obis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 231787..231880  
OTHER INFORMATION: exon O2 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 231870..231879  
OTHER INFORMATION: exon O1 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 234174..234321  
OTHER INFORMATION: exon O complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 237406..237428  
OTHER INFORMATION: exon Nbis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 239719..239807  
OTHER INFORMATION: exon N2 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 239719..239853  
OTHER INFORMATION: exon N complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240569  
OTHER INFORMATION: exon M1117 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240596  
OTHER INFORMATION: exon M1090 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240617  
OTHER INFORMATION: exon M1069 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240644  
OTHER INFORMATION: exon M52 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240824  
OTHER INFORMATION: exon M862 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240994  
OTHER INFORMATION: exon M692 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..241685  
OTHER INFORMATION: exon M1 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240800..240933  
OTHER INFORMATION: exon M51 complement g34872 gene  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 241686..243685  
OTHER INFORMATION: 5'regulatory region g34872 gene  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 290652..292652  
OTHER INFORMATION: 3'regulatory region g34665 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 292653..292841



OTHER INFORMATION: 8-293-130 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 206545  
OTHER INFORMATION: 8-292-198 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 207313  
OTHER INFORMATION: 8-251-322 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 208285  
OTHER INFORMATION: 8-289-322 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 208960  
OTHER INFORMATION: 8-287-249 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 209123  
OTHER INFORMATION: 8-287-86 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 209631  
OTHER INFORMATION: 8-285-319 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 210361  
OTHER INFORMATION: 8-283-278 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 210463  
OTHER INFORMATION: 8-283-176 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 210486  
OTHER INFORMATION: 8-283-153 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 210583  
OTHER INFORMATION: 8-283-56 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 210879  
OTHER INFORMATION: 8-282-345 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 210964  
OTHER INFORMATION: 8-282-260 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 210979  
OTHER INFORMATION: 8-282-245 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 211050  
OTHER INFORMATION: 8-282-174 : variable motif AAAGG or GAAGGAAGGAAGGAAGA  
NAME/KEY: allele  
LOCATION: 211132  
OTHER INFORMATION: 8-282-92 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 211247  
OTHER INFORMATION: 8-281-367 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 211315  
OTHER INFORMATION: 8-281-299 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 211366  
OTHER INFORMATION: 8-281-248 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 212520  
OTHER INFORMATION: 8-279-197 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 212821  
OTHER INFORMATION: 8-278-289 : polymorphic base C or T  
NAME/KEY: allele

Query Match 60.7%; Score 16.4; DB 4; Length 319608;  
Best Local Similarity 76.9%; Pred. No. 1.4e+02;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 ACCGGCGTTTAGGATCGACCTGA 27  
Db 120195 ACCCACTTTGAGGATCTACTCTGA 120170

RESULT 26

US-09-557-884-1  
; Sequence 1, Application US/09557884  
; Patent No. 6506581  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: The Nucleotide sequence of  
; the Haemophilus influenzae Rd Genome, Fragments  
; Thereof, and Uses Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3 1/2 inch diskette  
; COMPUTER: Dell Pentium  
; OPERATING SYSTEM: MS DOS v6.22  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/557,884  
; FILING DATE: 25-Apr-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/476,102  
; FILING DATE: JUN-5-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks  
; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: PB186P3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-309-8504  
; TELEFAX: 301-309-8439  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1830121 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-557-884-1  
Query Match 60.7%; Score 16.4; DB 4; Length 1830121;  
Best Local Similarity 94.4%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 8 GCTTTAGGATCGACCT 25  
Db 545007 GCTTTAGGATCGACCT 545024  
RESULT 27  
US-09-543-990A-1  
; Sequence 1, Application US/09643990A  
; Patent No. 6528289  
; GENERAL INFORMATION:  
; APPLICANT: Robert D. Fleischmann  
; Mark D. Adams  
; Owen White  
; Hamilton O. Smith  
; J. Craig Venter  
; TITLE OF INVENTION: The Nucleotide sequence of  
; the Haemophilus influenzae Rd Genome, Fragments  
; Thereof, and Uses Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville,  
; STATE: MD  
; COUNTRY: USA

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 2895:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1605 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...1605  
SEQUENCE DESCRIPTION: SEQ ID NO: 2895:  
US-09-107-532A-2895

Query Match 60.0%; Score 16.2; DB 4; Length 1605;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 CCGCGCGTTTAGGATCGACAC 23  
DB 1000 CCAGCGGTTAGATCGACAC 1020

RESULT 29  
US-09-252-991A-12862/c  
Sequence 12862, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 12862  
LENGTH: 405  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-12862

Query Match 59.3%; Score 16; DB 4; Length 405;  
Best Local Similarity 79.2%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACT 25  
DB 27 AGCGCGCTCTCGAGGATCGACACT 4

RESULT 30  
US-09-489-039A-888  
Sequence 888, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS v6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/643,990A  
FILING DATE: 23-AUG-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/487,429  
FILING DATE: 1995-06-07  
APPLICATION NUMBER: 08/426,787  
FILING DATE: 1995-04-21  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PE186P1C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-610-5790  
TELEFAX: 310-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-990A-1

Query Match 60.7%; Score 16.4; DB 4; Length 1830121;  
Best Local Similarity 94.4%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 GCTTTAGGATCGACACACT 25  
DB 545007 GCTTTAGGATCGACACT 545024

RESULT 28  
US-09-107-532A-2895  
Sequence 2895, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
CORRESPONDENCE ADDRESS:  
ADDRESS: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 888

; LENGTH: 651

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-888

Query Match

59.3%; Score 16; DB 4; Length 651;

Best Local Similarity 79.2%; Pred.No. 1.2e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACC 24

Db 97 CAGCGTCGTTTTATGATCGACACC 120

Search completed: June 20, 2004, 11:45:55

Job time : 48.9821 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 733.143 Seconds  
(without alignments)  
1596.226 Million cell updates/sec

Title: US-10-624-714-12  
Perfect score: 27  
Sequence: 1 caccgcgcttagatcgacactga 27

Scoring table:  
IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 300 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.em.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	27	100.0	17131	1	AE007080	AE007080 Mycobacte
C 2	27	100.0	306550	1	BX248342	BX248342 Mycobacte
C 3	27	100.0	348247	15	BX842579	BX842579 Mycobacte
C 4	19.4	71.9	139421	8	AC087545	AC087545 Oryza sat
C 5	19.4	71.9	300329	8	AE017097	AE017097 Oryza sat
C 6	18.6	70.4	151076	5	BX005194	BX005194 zebrafish
C 7	18.6	68.9	48	8	ATH552463	ATH552463 Arabidops
C 8	18.6	68.9	2000	6	AX509902	AX509902 Sequence
C 9	18.6	68.9	2851	1	AF548445	AF548445 Unculture
C 10	18.6	68.9	11481	1	AE014432	AE014432 Brucella
C 11	18.6	68.9	13916	1	AE009507	AE009507 Brucella
C 12	18.6	68.9	72598	8	AB007545	AB007545 Arabidops
C 13	18.6	68.9	111019	10	AL807391	AL807391 Mouse DNA
C 14	18.6	68.9	162646	9	AC078809	AC078809 Homo sapi
C 15	18	66.7	1287	6	AX122212	AX122212 Sequence
C 16	18	66.7	1287	6	BD164329	BD164329 Novel pol
C 17	18	66.7	1335	6	AX067039	AX067039 Sequence
C 18	18	66.7	1410	6	AX067037	AX067037 Sequence
C 19	18	66.7	1417	6	AX813887	AX813887 Sequence
C 20	18	66.7	11198	1	AE009079	AE009079 Agrobacte
C 21	18	66.7	14354	1	AE008044	AE008044 Agrobacte
C 22	18	66.7	84615	9	AL354747	AL354747 Human DNA
C 23	18	66.7	86257	9	AL954650	AL954650 Human DNA
C 24	18	66.7	109889	10	AL954830	AL954830 Mouse DNA
C 25	18	66.7	110000	2	LMFCHRI16_02	Continuation (3 of
C 26	18	66.7	110000	2	LMFCHRI16_03	Continuation (4 of
C 27	18	66.7	142340	10	AC124465	AC124465 Mus muscu
C 28	18	66.7	184424	2	AC106319	AC106319 Rattus no
C 29	18	66.7	206112	2	AC084412	AC084412 Mus muscu
C 30	18	66.7	213462	2	AC099282	AC099282 Rattus no
C 31	18	66.7	213474	2	AC103199	AC103199 Rattus no
C 32	18	66.7	223040	2	AC128102	AC128102 Rattus no
C 33	18	66.7	231274	2	AC099393	AC099393 Rattus no
C 34	18	66.7	231509	2	AC131426	AC131426 Rattus no
C 35	18	66.7	232415	2	AC095640	AC095640 Rattus no
C 36	18	66.7	237378	2	AC087038	AC087038 Mus muscu
C 37	18	66.7	246108	2	AC103535	AC103535 Rattus no
C 38	18	66.7	255746	2	AC122622	AC122622 Rattus no
C 39	18	66.7	323450	1	SM5591790	SM5591790 Sinorhizo
C 40	18	66.7	337200	1	AP005280	AP005280 Corynebac
C 41	18	66.7	338579	1	AP003004	AP003004 Mesorhizo
C 42	18	66.7	349980	6	AX127149	AX127149 Sequence
C 43	17.8	65.9	2319	1	BSORF4	X74289 B.stearothe
C 44	17.8	65.9	4839	1	NCSE6	Y08751 B.stearothe
C 45	17.8	65.9	119882	8	NCSE6	AL670004 Neurospor
C 46	17.8	65.9	301550	1	AP005079	AP005079 Vibrio pa
C 47	17.6	65.2	894	1	AF146711	AF146711 Chromobac
C 48	17.6	65.2	1207	1	CHRPHERYD	M55915 Chromobacte
C 49	17.6	65.2	2407	8	AK069904	AK069904 Oryza sat
C 50	17.6	65.2	4382	9	HSN802264	AL137724 Homo sapi
C 51	17.6	65.2	4919	6	AX405811	AX405811 Sequence
C 52	17.6	65.2	5135	1	TTNUSAINF	Z48001 T.thermophi
C 53	17.6	65.2	121796	2	AC147335	AC147335 Pan trogl
C 54	17.6	65.2	154737	9	AC078834	AC078834 Homo sapi
C 55	17.6	65.2	171452	2	AC122297	AC122297 Mus muscu
C 56	17.6	65.2	190220	5	AL935305	AL935305 zebrafish
C 57	17.6	65.2	241213	2	AC139952	AC139952 Rattus no
C 58	17.6	65.2	249406	2	AC118963	AC118963 Rattus no
C 59	17.6	65.2	256017	2	AC126312	AC126312 Rattus no
C 60	17.6	65.2	298600	1	AP005959	AP005959 Bradyrhiz
C 61	17.6	65.2	302998	1	AE016921	AE016921 Chromobac
C 62	17.6	65.2	306214	2	AC123205	AC123205 Rattus no
C 63	17.6	65.2	325483	1	AP005050	AP005050 Streptomy
C 64	17.6	65.2	340900	1	SM5591791	AL591791 Sinorhizo
C 65	17.4	64.4	538	1	MTU85630	U85630 Mycobacteri





C 212	17	63.0	186566	2	BX5111265	Danio rer	BX5111265	Danio rer	285	16.8	62.2	260894	2	AC119873	Mus muscu
C 213	17	63.0	186566	2	BX5111265	Danio rer	BX5111265	Danio rer	286	16.8	62.2	261151	2	AC096326	Rattus no
C 214	17	63.0	193317	9	AC098645	Papio anu	AC098645	Papio anu	C 287	16.8	62.2	279252	2	AC093941	Rattus no
C 215	17	63.0	196238	2	AC132286	Mus muscu	AC132286	Mus muscu	C 288	16.8	62.2	310850	1	AP005142	Streptoco
C 216	17	63.0	197581	2	AC137043	Rattus no	AC137043	Rattus no	C 289	16.6	61.5	546	6	AX461531	Sequence
C 217	17	63.0	203050	1	AL646071	Ralstonia	AL646071	Ralstonia	C 290	16.6	61.5	546	6	AX505992	Sequence
C 218	17	63.0	203830	2	AC096894	Rattus no	AC096894	Rattus no	C 291	16.6	61.5	546	8	AY081728	Arabidops
C 219	17	63.0	207823	10	AL929142	Mouse DNA	AL929142	Mouse DNA	C 292	16.6	61.5	546	8	AY125540	Arabidops
C 220	17	63.0	212779	9	AC104031	Homo sapi	AC104031	Homo sapi	C 293	16.6	61.5	581	8	LLOC010974	Lilium lo
C 221	17	63.0	213119	10	AC091694	Mus muscu	AC091694	Mus muscu	C 294	16.6	61.5	587	6	BD073325	Novel nuc
C 222	17	63.0	213505	2	AC132571	Mus muscu	AC132571	Mus muscu	C 295	16.6	61.5	587	8	AF090826	Lilium lo
C 223	17	63.0	213970	2	AC097765	Rattus no	AC097765	Rattus no	C 296	16.6	61.5	588	8	AB003781	Lilium lo
C 224	17	63.0	214425	2	AC136109	Rattus no	AC136109	Rattus no	C 297	16.6	61.5	601	8	AB003782	Lilium lo
C 225	17	63.0	215987	2	AC134464	Mus muscu	AC134464	Mus muscu	C 298	16.6	61.5	1423	8	AF385706	Arabidops
C 226	17	63.0	217768	10	AC074336	Mus muscu	AC074336	Mus muscu	C 299	16.6	61.5	1444	8	AY060561	Arabidops
C 227	17	63.0	218567	2	AC107735	Mus muscu	AC107735	Mus muscu	C 300	16.6	61.5	1587	8	BT009450	Triticum
C 228	17	63.0	219107	2	AC141343	Rattus no	AC141343	Rattus no							
C 229	17	63.0	222528	2	AC107118	Rattus no	AC107118	Rattus no							
C 230	17	63.0	225079	2	AC126203	Rattus no	AC126203	Rattus no							
C 231	17	63.0	226832	2	AC110969	Rattus no	AC110969	Rattus no							
C 232	17	63.0	227709	2	BX571722	Danio rer	BX571722	Danio rer							
C 233	17	63.0	237343	2	BX569792	Danio rer	BX569792	Danio rer							
C 234	17	63.0	239377	2	BX511167	Danio rer	BX511167	Danio rer							
C 235	17	63.0	239377	2	BX511167	Danio rer	BX511167	Danio rer							
C 236	17	63.0	240886	2	AC096519	Rattus no	AC096519	Rattus no							
C 237	17	63.0	245358	2	AC136008	Mus muscu	AC136008	Mus muscu							
C 238	17	63.0	247012	2	AC109008	Rattus no	AC109008	Rattus no							
C 239	17	63.0	247381	2	AC134193	Rattus no	AC134193	Rattus no							
C 240	17	63.0	247381	2	AC106199	Rattus no	AC106199	Rattus no							
C 241	17	63.0	251970	2	AC094791	Rattus no	AC094791	Rattus no							
C 242	17	63.0	252602	2	AC091704	Mus muscu	AC091704	Mus muscu							
C 243	17	63.0	258276	2	AC123659	Rattus no	AC123659	Rattus no							
C 244	17	63.0	261258	2	AC091365	Rattus no	AC091365	Rattus no							
C 245	17	63.0	264321	10	AL807762	Mouse DNA	AL807762	Mouse DNA							
C 246	17	63.0	266456	2	AC120829	Rattus no	AC120829	Rattus no							
C 247	17	63.0	279012	2	AC126134	Rattus no	AC126134	Rattus no							
C 248	17	63.0	280277	2	AC125932	Rattus no	AC125932	Rattus no							
C 249	17	63.0	287086	2	BX530407	Danio rer	BX530407	Danio rer							
C 250	17	63.0	287086	2	BX530407	Danio rer	BX530407	Danio rer							
C 251	17	63.0	289671	2	AC097291	Rattus no	AC097291	Rattus no							
C 252	17	63.0	296131	2	AL928718	Mus muscu	AL928718	Mus muscu							
C 253	17	63.0	300094	2	AC113682	Rattus no	AC113682	Rattus no							
C 254	17	63.0	303698	3	AB003434	Drosophil	AB003434	Drosophil							
C 255	17	63.0	304517	1	AB016789	Pseudomon	AB016789	Pseudomon							
C 256	17	63.0	308760	2	AC126721	Rattus no	AC126721	Rattus no							
C 257	17	63.0	310539	2	AC118854	Rattus no	AC118854	Rattus no							
C 258	17	63.0	310539	2	AC118854	Rattus no	AC118854	Rattus no							
C 259	17	63.0	316050	1	BX321859	Nitrosomo	BX321859	Nitrosomo							
C 260	17	63.0	325667	2	AC112413	Rattus no	AC112413	Rattus no							
C 261	17	63.0	329100	1	SME591787	Sinorhizo	AL591787	Sinorhizo							
C 262	17	63.0	332551	2	AC094116	Rattus no	AC094116	Rattus no							
C 263	17	63.0	333195	2	AC126635	Rattus no	AC126635	Rattus no							
C 264	17	63.0	333878	2	AC111474	Rattus no	AC111474	Rattus no							
C 265	17	63.0	339391	2	AC117149	Rattus no	AC117149	Rattus no							
C 266	17	63.0	345345	2	AC126167	Rattus no	AC126167	Rattus no							
C 267	17	63.0	348942	1	BX572596	Rhodopseu	BX572596	Rhodopseu							
C 268	17	63.0	349970	1	BX571659	Wolirella	BX571659	Wolirella							
C 269	16.8	62.2	804	1	AB039546	Pseudomon	AB039546	Pseudomon							
C 270	16.8	62.2	1146	1	BSAMAA	Y08753	B.stearothe	Y08753							
C 271	16.8	62.2	3625	8	BDNA28RR	X50411	B.emersonii	X50411							
C 272	16.8	62.2	51385	1	AB014162	Streptoco	AB014162	Streptoco							
C 273	16.8	62.2	57876	9	AC007074	Homo sapi	AC007074	Homo sapi							
C 274	16.8	62.2	62736	2	AC103774	Mus muscu	AC103774	Mus muscu							
C 275	16.8	62.2	112513	2	AC141825	Apis mell	AC141825	Apis mell							
C 276	16.8	62.2	146258	2	AC023479	Homo sapi	AC023479	Homo sapi							
C 277	16.8	62.2	148911	2	AL353142	Homo sapi	AL353142	Homo sapi							
C 278	16.8	62.2	160012	9	AL136231	Human DNA	AL136231	Human DNA							
C 279	16.8	62.2	176977	2	AC105784	Rattus no	AC105784	Rattus no							
C 280	16.8	62.2	206801	10	AC073761	Mus muscu	AC073761	Mus muscu							
C 281	16.8	62.2	217323	2	AC115724	Mus muscu	AC115724	Mus muscu							
C 282	16.8	62.2	217788	2	AC124577	Mus muscu	AC124577	Mus muscu							
C 283	16.8	62.2	237565	2	AC134023	Rattus no	AC134023	Rattus no							
C 284	16.8	62.2	242966	2	AC111931	Rattus no	AC111931	Rattus no							

ALIGNMENTS

RESULT 1  
LOCUS AE007080/c  
DEFINITION Mycobacterium tuberculosis CDC1551, section 166 of 280 of the complete genome.  
ACCESSION AE007080  
VERSION AE007080.1  
KEYWORDS GI:13882094  
SOURCE Mycobacterium tuberculosis CDC1551  
ORGANISM Mycobacterium tuberculosis CDC1551  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.  
REFERENCE 1 (bases 1 to 17131)  
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.  
TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 17131)  
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
Location/Qualifiers  
1..17131  
/organism="Mycobacterium tuberculosis CDC1551"  
/mol\_type="genomic DNA"  
/strain="CDC1551"  
/db\_xref="taxon:83331"  
/note="clinical strain"  
104..1357  
/gene="WT2390"  
/gene="WT2390"  
104..1357  
/note="similar to SP:P71884 PID:1524272 GB:AL123456; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="PE family protein"  
/protein\_id="AAK4682.1"  
/db\_xref="GI:13882095"  
/translation="MTANALARKDRHFTTINVPSSLDNRRSGRTSMOFLSVPE  
QVESAQDLGIRLSASAAAGPTTAVUSAEDVSTASTAFGAGRCOVLISA  
QASAFHDFVNLKTCATYRNTFANQSNVLNAPARSLLGHPSAESVQNSAP



```

/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46691.1"

Query Match      100.0%; Score 27; DB 1; Length 17131;
Best Local Similarity 100.0%; Pred. NO. 0.024;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTGA 27
    |||||
Db 3777 CACCGCGCTTTAGGATCGACACCTGA 3751

RESULT 2
BX248342/c 306550 bp DNA linear BCT 11-JUN-2003
LOCUS Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment
DEFINITION 9/14
ACCESSION BX248342 BX248333
VERSION BX248342.1 GI:31619031
KEYWORDS complete genome.
SOURCE Mycobacterium bovis subsp. bovis AF2122/97
ORGANISM Mycobacterium bovis subsp. bovis AF2122/97
Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE 1
AUTHORS Garnier,T., Eiglmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
Pryor,M., Duthoy,S., Grondin,S., Lacroix,C., Monsemp,C., Simon,S.,
Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,
Wheeler,P.R., Parkhill,J., Barrall,B.G., Cole,S.T., Gordon,S.V. and
Hewinson,G.
TITLE The complete genome sequence of Mycobacterium bovis
JOURNAL Online Publication
REMARK PNAS 10.1073/pnas.1130426100 ( Microbiology )
REFERENCE 2 (bases 1 to 306550)
AUTHORS Garnier,T.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
Bacterienne Institut Pasteur 28,rue du Dr Roux 75724 PARIS cedex
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
Laboratories Agency Weybridge, Woodham Lane, New Haw Addlesstone,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK. PF4 Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
rue du Docteur Roux, 75724 Paris Cedex 15, France
Location/Qualifiers
1..306550
/organism="Mycobacterium bovis subsp. bovis AF2122/97"
/mol_type="genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"
201..1286
/gene="adhE2"
/locus_tag="Mb2283"
201..1286
/gene="adhE2"
/locus_tag="Mb2283"
/EC_number="1.2.1.-"
/Note="Mb2283, adhE2, len: 361 aa. Equivalent to Rv2259,
len: 361 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 361 aa overlap). Probable adhE2,
zinc-containing alcohol dehydrogenase, similar to several,
especially mycothiol-dependent formaldehyde dehydrogenase
from Amycolatopsis methanolica P80094 (360 aa). Contains
P800059 Zinc-containing alcohol dehydrogenases signature.
FASTA scores: >sp|P80094|FADH_AMEYE
NAD/MYCOTHOL-DEPENDENT FORMALDEHYDE DEHYDROGENASE
(MD-FALDH) Length = 360, Expect = e-156, Identities =
268/358 (74%). Also similar to Rv0162c, (MTC128.02c, 35.0%
identity in 371 aa overlap)."
/codon_start=1
/transl_table=11
/product="Probable zinc-dependent alcohol dehydrogenase
AdhE2"
/protein_id="CAD97144.1"
/db_xref="GI:31619032"
/transl_table=11
/product="Probable zinc-dependent alcohol dehydrogenase
AdhE2"
/protein_id="CAD97144.1"
/translation="MSQTVRGVIARQKGPVELVNI VVPDPGAEAVDVVTCGVCHT
DLTVREGGINDEYFPLGLGHEAAGLIEAVGPGVTAVEGDFVILNWEAVCCGCRCKRG
RPRVCFDTEAEOKMTLTDGTELTALGICAFADKTLVHSGQCTKVDPAADPAVAGLL
CGVMAGLGAATNGVTRDDTVAVICGGVGDAIAGAALVAGAKRIIAVDTDTKLD
WARTFGATHVNAREVDDVQAI GGLDTGFGADVVIDAVGRPETYQQAIFYARDIAGTVV
LVGPTDMDLMDPLVDFSGGALKSSWYGDCLPESDFTLLDLVQGRPLQRFVYS
ERIGLEDVEAFHMGKGKVLRSVWML"
1286..1921
/locus_tag="Mb2284"
1286..1921
/locus_tag="Mb2284"
/Note="Mb2284, -, len: 211 aa. Equivalent to Rv2260, len:
211 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 211 aa overlap). Conserved
hypothetical protein, similar to hypothetical proteins
Rv0634c, Rv1637c, Rv3677c, Rv2581c from Mycobacterium
tuberculosis and to various hydrolases. FASTA scores:
sp|O6154|O6154_HYPOTHETICAL_21.3_KD_PROTEIN (200 aa)
opt: 355, E(): 4e-15; (37.4% identity in 198 aa
overlap)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD97145.1"
/db_xref="GI:31619033"
/translation="MAAIEKVITHGTFFDGGSEVNDNINLVGDSEVVVFDAAHHA
APIIDVAGGRKVVAVICTGHNDHVTVAPELGTALDPAVLMHFGDAVLRMTFDPKSF
RAVSDGDVAVRGVGTETLRHTPHGSPGVCWYAPELGPGTGVTFSGDTLPAAGPGATG
RSYSDPTILRSISGRGLGALPGDTVVHTGHGDDTTIGDEIVHYEEVWARGH"
complement (1998..3506)
/locus_tag="Mb2285c"
complement (1998..3506)
/locus_tag="Mb2285c"
/Note="Mb2285c, -, len: 502 aa. Equivalent to Rv2262c and
Rv2261c, len: 360 aa and 140 aa, from Mycobacterium
tuberculosis strain H37Rv, (94.7% identity in 357 aa
overlap and 100.0% identity in 140 aa overlap). Conserved
hypothetical protein, with function unknown but some
similarity to N-terminal 70% of
P23930|P77703|LNT_ECOLI|CUTE|B0657 APOLIPOPROTEIN
N-ACYLTRANSFERASE (EC 2.3.1.-) from Escherichia coli
strain K12 (512 aa), FASTA scores: opt: 239, E(): 1.6e-07,
(30.4% identity in 359 aa overlap). Note that neighboring
ORF shows similarity to N-terminal part of PCC6803
apolioprotein N-acyltransferase from Synecocystis sp.,
suggesting possibility of frameshift. Sequence of clones
from two sources has been checked but no error found.
Appear to be two extra bases at position 1876970 compared
to CDG1551 strain. Conserved hypothetical protein, with
function unknown but some similarity to C-terminal end of
PCC6803 apolioprotein N-acyltransferase from
Synecocystis sp. Note that next ORF shows similarity to
N-terminal part of P74055 APOLIPOPROTEIN N-ACYLTRANSFERASE
from Escherichia coli (519 aa), FASTA scores: opt: 142,
E(): 0.007, (29.9% identity in 117 aa overlap), suggesting
possible frameshift. Sequence of clones from two sources
has been checked but no error found.
REMARK-M. bovis-M. tuberculosis: In Mycobacterium
tuberculosis strain H37Rv, Rv2262c and Rv2261c exist as 2
genes. In Mycobacterium bovis, a 2 bp deletion (ct-*)
results in a single product which is more similar to
Rv2262c."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD97146.1"

```



FT penicillin-binding membrane protein, similar to many  
FT bacterial PBP2 proteins e.g.  
FT P11882|PBP2\_NEIME|PENA|NNA2072|NMB0413|penicillin-binding  
FT protein 2 (pbp-2) from Neisseria meningitidis (serogroups A  
FT and B) (581 aa), FASTA scores: Opt: 665, E(): 1.6e-31,  
FT (33.2% identity in 591 aa overlap); etc. Also similar to  
FT P20016c and Rv2864c from Mycobacterium tuberculosis  
FT (Rv08-10). Contains F500017 possible ATP/GTP-binding site  
FT motif A (P-loop) near C-terminus. FASTA best: PBP2\_NEIME  
FT P11882|penicillin-binding protein 2 (pbp-2). (581 aa) opt:  
FT 665, E(): 1.6e-31; (33.2% identity in 591 aa overlap)."  
FT /transl\_table=11  
FT /function="Involved in peptidoglycan biosynthesis."  
FT /gene="pbpB"  
FT /locus\_tag="Rv2163c"  
FT /product="Probable penicillin-binding membrane protein  
FT pbpB"  
FT /protein\_id="CAB08664.1"  
FT /translation="MSRAAPRASQOSTPRARGLRPPGGAQVEGQKRGKTKAKQA  
FT QEAACKSRDADVAPAGSTEARTRQVDVGTGASFFVPHRTGNVILVLMVAAT  
FT QLFQVSHAGLRQAQAGLKVDVQPAAGSIVDRNDLAFITIGARLTFQFKIR  
FT QLEBAARKTGAADPQORLDIAQEAQVAGLNKPDAAALVKLQSDTFVYLARVDP  
FT AVASAKYFEGVAGRDQLRCYQGSLLAANVGGIDWDGHLGLSDSLAVLAGDVG  
FT SVYDRGSDGVVPGSYRNRKAVHGSTVLTLDNDIQFYVQQQAKNLGSAHNVSA  
FT VLLDAKTEVLAMANDNTFDSQDIGROGDKLGNPAVSPFEPGSVKNIVAAAGVIEH  
FT GLSSDVEVLQGVGTQMGVTVHAWEGHVMETTTGVGKSSNVGTLMSQVGPERY  
FT YDMLRKFGLGORTGVLGPGESAGVPPIDQSGSTFANLPIGQGLSLTQMOTMYQAI  
FT ANDGVVPPRIKATVADGRTSEPRDDRVVSAQTAQIVRQMLRAVQVORDPMGQIQ  
FT GTGTVAGVFGQMAKTAQOINFGCGYDFDVITTFAGIATADNPRVYIGIMLNDP  
FT ARNSDGAAGSAALPHNIAQMLQRENVPISPDGPPPLVQAT"  
FT complement(470. .493)  
FT /note="P500017 ATP/GTP-binding site motif A (P-loop)"  
FT complement(12245. .3399)  
FT /note="Rv2164c. (MTCY270.04), len: 384 aa. Probable pro-  
FT rich conserved membrane protein, equivalent to  
FT M0907|AL022602 putative conserved membrane protein from  
FT Mycobacterium leprae (377 aa) (AL022602), FASTA scores:  
FT Opt: 1495, E(): 1.7e-56, (62.217% identity in 397 aa  
FT overlap)."  
FT /transl\_table=11  
FT /function="UNKNOWN"  
FT /locus\_tag="Rv2164c"  
FT /product="PROBABLE CONSERVED PROLINE RICH MEMBRANE PROTEIN"  
FT /translation="WRAKREAPKSSRRRRADSPAAATRTTNSAPRRRISRAGK  
FT TSAPQRAVSRRPGQSPMLSPDRPAPAKNTSOAKARAKAKAPKLVPPMERL  
FT ARLTSLDLRPLANKVPFVVLVIGSLGVLGLTLMSTDAERSYQLSNARERTML  
FT QCKEALRDVREASAPALAEARQGMIFRTDLHLVQDPGNVYVVPKRPADGVP  
FT PPLNTKLPEPPPPPPKPAVLEVPVRVTPGDDPAPPAKSPSEVILTPDGTATGG  
FT ATHLPQTQGPQGPVTPGAPMPAPPLGAPSPAPAPENVPVLOVGAAPAGLPGPA  
FT PVAATPGLSGSQMPVAPPVPPVPPGQFGVTPVPTAPGAPR"  
FT complement(3396. .4586)  
FT /note="Rv2165c. (MTCY270.03), len: 396 aa. Conserved  
FT hypothetical protein, shows strong similarity to several  
FT hypothetical bacterial proteins but has extra 80 aa  
FT residues at N-terminus FASTA best: YLXA\_BACSU Q07876  
FT hypothetical 35.3 kDa protein in ftsl (311 aa) opt: 781,  
FT E(): 0; (45.6% identity in 296 aa overlap), BELONGS TO THE  
FT YABC (E. COLI), YLXA (B. SUBTILIS) FAMILY"  
FT /transl\_table=11  
FT /function="UNKNOWN"  
FT /locus\_tag="Rv2165c"  
FT /product="CONSERVED HYPOTHETICAL PROTEIN"  
FT /protein\_id="CAB08662.1"  
FT /translation="WQTRAPWSLPEATLAVPPNARFVSSRDGLGAGAPGIAASRSTAC  
FT QTWGITVADPGSGPTGFGHVPVLAORCFELLTPALTRYVPDGSQAVLLDATTGAGHAG  
FT BRFEGLGLSLIGLDRTDPAVDALVARSRLVFAEDRLTVHTRVDCLGAARSYARVG  
FT SYGILEDLVSSQDLREAERGFAVADALDMDMTTBLTAADIWNTYDREAALADIL  
FT RYGEERFARRIKVRRRATFTSTAEVLVLLYQAIIPAPRVRGHHFAKRTFQALR  
FT IAVNDELESRTVPAALDAIGRIATVAYQSLDRIIVKRVFAEVAAPAGPVE  
FT LPGHEPRFSITGHAERASVAEIERNPRSTFVRLRALQVRHRAQSQWATERKGS"  
FT complement(4588. .5019)  
FT /note="Rv2166c. (MTCY270.02), len: 143 aa. Conserved

FT hypothetical protein; shows strong similarity to several  
FT hypothetical bacterial proteins such as YLLB\_BACSU P55343.  
FT Is equivalent to Mycobacterium leprae hypothetical protein  
FT MLO905 (143 aa, 92% identity) MLCB268.11c  
FT >sp|O69561|YL66 MYCLE HYPOTHETICAL 16.1 KDA PROTEIN MLO905  
FT >gi|3080482|emb|CAA18677.1|(AL022602)  
FT >gi|13092975|emb|CAC31286.1|(AL583920). FASTA scores:  
FT MLO905|MLO905 conserved hypothetical protein (143 aa) opt:  
FT 873, E(): 3.1e-52; 92.254% identity in 142 aa overlap;  
FT YLLB\_BACSU P55343 hypothetical 16.6 kDa protein (143 aa)  
FT opt: 340, E(): 3.6e-17; (35.0% identity in 143 aa overlap).  
FT BELONGS TO THE YABB (E. COLI), YLLB (B. SUBTILIS), MG221 (E.  
FT M. GENTILIIUM) FAMILY"  
FT /transl\_table=11  
FT /function="UNKNOWN"  
FT /locus\_tag="Rv2166c"  
FT /product="CONSERVED HYPOTHETICAL PROTEIN"  
FT /protein\_id="CAB08661.1"  
FT /translation="MFLGTYTPKLDKDKBLTIPAKFPDALAGLMVTKSQDHSLVYPR  
FT AAFEQARASAKAPRNPAPARAPLRNLAAGTDECHDPSQRTILSDHRRYASLSKDCV  
FT VICAVYLIRWQAQWQNTQQIHEENFSAASDEALGDIIF"  
FT complement(5278. .5305)  
FT /note="28 bp Inverted repeat at the left end of IS6110;  
FT GAGTCTCCGACTCACCGGCGGCTTCA"  
FT complement(5278. .6632)  
FT /note="IS6110-6. len: 1355 bp. Insertion sequence IS6110."  
FT /insertion\_seq="IS6110-6"  
FT complement(5320. .56360)  
FT /notes="Rv2167c. (MTCY270.01), len: 346 aa. Probable IS6110  
FT transposase. FASTA best: TRA9 MYCTU P19774 putative  
FT transposase for insertion sequence (identical)."  
FT /transl\_table=11  
FT /function="REQUIRED FOR THE TRANSPOSITION OF THE INSERTION  
FT ELEMENT IS6110."  
FT /locus\_tag="Rv2167c"  
FT /product="PROBABLE TRANSPOSASE"  
FT /protein\_id="CAA17494.1"  
FT /translation="AALAAAGORRIAGKGRDFKRVGLRGARPASTLITRPIADHOG  
FT HREGPDGLRWGVESICTOLTGLGVPITPSTYVDHINRPSRRELRLGELKHSRVHAA  
FT NYGVGARKVWLTLNREGLEVARTVERLMTKLGLSGTTRGKARTITADPATAPADL  
FT VQRFGPPAPNRLVADLIYVWAGVFAIVFVDAIYARRILGWATWSYMDLDAI  
FT EGAIRWTRQEGVLDLKDVIHTRDGSQYTSIRFSLRABAGIQPSVGAVGSSYDNLAE  
FT TNLGYLTLLIKLPGKWPWSIEDVELATARVDFWDFNHRRLYQYCGDVPVPELEAAVYQAQ  
FT QRPAAQ"  
FT complement(6255. .6581)  
FT /note="Rv2168c. (MTV021.01c), len: 108 aa. Probable IS6110  
FT transposase. FASTA scores: O08155|O08155 HYPOTHETICAL 12.0  
FT kDa PROTEIN (108 aa) opt: 697, E(): 0; (100.0% identity in  
FT 108 aa overlap). TBPase score is 0.928."  
FT /transl\_table=11  
FT /function="REQUIRED FOR THE TRANSPOSITION OF THE INSERTION  
FT ELEMENT IS6110."  
FT /locus\_tag="Rv2168c"  
FT /product="PROBABLE TRANSPOSASE"  
FT /protein\_id="CAA17472.1"  
FT /translation="MSGSSRRYPPPELRERAYRMVAEIRGQHDSEWAAISEVARLLGVG  
FT CAETVKWVRQAQVDAGAPFGTTTBESAEKLRLRDNALRRANILKTASAFFAELD  
FT RPAP"  
FT complement(6605. .6632)  
FT /note="28 bp Inverted repeat at the right end of IS6110,  
FT TGAACCCGCCGCGATGCCGAGATC"  
FT complement(6726. .7130)  
FT /evidence=EXPERIMENTAL

Query Match 100.0%; Score 27; DB 15; Length 348247;

Best Local Similarity 100.0%; Pred No. 0.023; Mismatches 0; Gaps 0;

Matches 27; Conservative 0; Indels 0; Gaps 0;

Qy 1 CACCCGCGTTTAGGATCGACACCTGA 27

Db 179460 CACCCGCGTTTAGGATCGACACCTGA 179434

```

RESULT 4
AC087545
LOCUS
DEFINITION
    Orzya sativa (japonica cultivar-group) linear PLN 30-MAY-2002
    nbx0019M20, complete sequence.
ACCESSION
    AC087545
VERSION
    AC087545.3 GI:21263269
KEYWORDS
    HTG.
SOURCE
    Orzya sativa (japonica cultivar-group)
    Orzya sativa (japonica cultivar-group)
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
    Ehrhartoideae; Oryzeae; Oryza.
    1 (bases 1 to 139421)
REFERENCE
    AUTHORS
    Llaica, V., Song, R., Young, S., Kavchok, S., Ward, K. and Messing, J.
    TITLE
    Orzya sativa (japonica cultivar-group) chromosome 10 clone
    nbx0019M20
    JOURNAL
    Unpublished
    AUTHORS
    2 (bases 1 to 139421)
    Llaica, V., Young, S., Kavchok, S., Clark, T., Charydczac, G., Choi, V.,
    Nevill-Manning, C. and Messing, J.
    TITLE
    Direct Submission
    JOURNAL
    Submitted (08-JAN-2001) The Plant Genome Initiative at Rutgers -
    Waksman Institute, Rutgers University, 190 Frelinghuysen Road,
    Piscataway, NJ 08873
    Chromosome 10
    REMARK
    3 (bases 1 to 139421)
    Llaica, V., Young, S., Kavchok, S., Ward, K., Charydczac, G. and
    Messing, J.
    TITLE
    Direct Submission
    JOURNAL
    Submitted (02-MAR-2001) The Plant Genome Initiative at Rutgers -
    Waksman Institute, Rutgers University, 190 Frelinghuysen Road,
    Piscataway, NJ 08873
    Chromosome 10
    REMARK
    4 (bases 1 to 139421)
    Llaica, V., Song, R., Young, S., Kavchok, S., Ward, K. and Messing, J.
    TITLE
    Direct Submission
    JOURNAL
    Submitted (30-MAY-2002) The Plant Genome Initiative at Rutgers -
    Waksman Institute, Rutgers University, 190 Frelinghuysen Road,
    Piscataway, NJ 08873, USA
    On May 30, 2002 this sequence version replaced gi:13184927.
    This BAC overlaps with rice BAC nbx0073D04 (AC087542) and
    nbx0012D19 (AC087546).
FEATURES
    source
    1..139421
    /organism="Orzya sativa (japonica cultivar-group)"
    /mol_type="genomic DNA"
    /sub_strains="japonica"
    /cultivar="Nipponbare"
    /db_xref="taxon:39947"
    /chromosome="10"
    /clone="nbx0019M20"
ORIGIN
    Query Match 71.9%; Score 19.4; DB 8; Length 139421;
    Best Local Similarity 95.2%; Pred. No. 1.5e+02;
    Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
    YQ 6 GGCCTTTAGGATCGACACCTG 26
    Db 18797 GCACCTTTAGGATCGACACCTG 18817
RESULT 5
AE017097
LOCUS
DEFINITION
    Orzya sativa (japonica cultivar-group) linear PLN 06-JUN-2003
    77 of the complete sequence.
ACCESSION
    AE017097
VERSION
    AE017097.1 GI:31432234
KEYWORDS
    Orzya sativa (japonica cultivar-group)
    Orzya sativa (japonica cultivar-group)

```

```

Bakaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 300029)
The Rice Chromosome 10 Sequencing Consortium
In-depth view of structure, activity, and evolution of rice
chromosome 10
Science 300, 1566-1569 (2003)
2 (bases 1 to 300029)
Buell, C.R., Wing, R.A., McCombie, W.R., Messing, J. and Yuan, Q.
Direct Submission
Submitted (05-MAY-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
This is the pseudomolecule for rice chromosome 10, which was
constructed by resolving discrepancies between overlapping BACs,
trimming the overlap regions, and linking the unique sequences to
form a contiguous sequence. Genes in individual BAC clone were
identified by a combination of several methods: Gene prediction
programs, searches of the complete sequence against a peptide
database and EST databases. Genes with similarity to other proteins
are named after the database hits. Genes without significant
peptide similarity but with EST similarity are named as unknown
proteins. Genes without protein or EST similarity, that are
predicted by more than two gene prediction programs over most of
their length are annotated as hypothetical proteins. Genes
encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/addy/tRNAscan-SE/).
Location/Qualifiers
1..300029
/organism="Orzya sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="10"
misc_feature
<1..47797
/notes="Chromosome Sequence Derivation: nucleotide sequence
in this region was derived from BAC clone nbx0073D04
(GB:AC087542)."
<3344..>4619
/locus_tag="OSJNBA0073D04.33"
/notes="Predicted by fgenesh"
join(<3344..3349,4158.>4619)
/locus_tag="OSJNBA0073D04.33"
join(3344..3349,4158..4619)
/locus_tag="OSJNBA0073D04.33"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAP53890.1"
/db_xref="GI:31432235"
/translation="MEIDDKGRPMSEBELTEKHANAVQIAKWPPEPVGVAKVNIID
GFRKETGDACAGIIVRDCRGLVLAACKLPRCSATQAEALACELGVELATNWHMP
IILESDNADVAGLITQASRAEWGIIAEIRVAMQCLIQVQHVHKRDSNRG"
<6000..>6671
/locus_tag="OSJNBA0073D04.34"
/notes="Predicted by fgenesh"
<6000..>6671
/locus_tag="OSJNBA0073D04.34"
6000..6671
/locus_tag="OSJNBA0073D04.34"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAP53891.1"
/db_xref="GI:31432236"
/translation="MGASASSGERRRRRCGSAIIVADTVSEHFDLTIDGYSHIKAMLP
TGKATCSPTPTAGCHEWCVDFYPNGKAAAGADMIQFFLRRAAKAKAAAVVQ
VRFDLLGSGSPQAGSGTGVLSFVDPVQSGSWIRRDFAEQACGKDDFTVRFA
VTVFRCRTAAAPESRPRRPPSPASPLPMCSYTCVTDCQTRNVRSAFTRPSSSTD
GGCQV"
complement(<7340..>10632)
/locus_tag="OSJNBA0073D04.1"
/notes="similar to cinnamyl alcohol dehydrogenase
GB:AA19536.GI:19849248 (Lolium perenne); EST C72961,

```

mRNA  
C24898, BM421887 from this gene"  
complement [join(<7340, .7779,7864, .8531,8806, .8919,  
10384, .>10632)]  
/locus\_tag="OSJNBa0073D04.1"  
complement [join(7568, .7779,7864, .8531,8806, .8919,  
10364, .>10632)]  
/locus\_tag="OSJNBa0073D04.1"  
/codon\_start=1  
/product="putative cinnamyl alcohol dehydrogenase"  
/translation="MATAEAAAARAGGGGSGNRGEASDAGRRQRRRAAGARTAG  
GCEAAAARAGGGGSGNRGEASDAGRRQRRRAAGARTAGGGGSGNRGEASDAGRRQRRRAAGARTAG  
FCGICHSGLHCIGKNEWHYIYPLVGEHIACTVTEGVKNVTRFKAGDRVGVGCMWNSC  
RSCSCNNGFENHCPGEGVFTVNSVDKGTVTYGGYSVMVYVHERFVYVPEAMELDVG  
APLICAGITVTPKHYGLNAPKGVHGLGGLGHVAVFAFAFGKLVVIVSSPGK  
KREALERIGADVFVSSAEAREASMTDGVINTVSANTPMAPYLALLKPVNEMILV  
GLPENLEVPFSLVHGNRTLAGNIGMADTQEMIILAAKHGVTADIEVIGADVNT  
AMERLAKADVRYFRVIDVGNTHAAAAE"  
/locus\_tag="OSJNBa0073D04.36"  
/note="similar to retinitis pigmentosa GTPase regulator  
GB:AG00551 GI:9937379 (Homo sapiens)"  
join(<14775, .15042,15172, .15370,15464, .15728,  
15842, .>15934)  
/locus\_tag="OSJNBa0073D04.36"  
join(14775, .15042,15172, .15370,15464, .15728,15842, .>15934)  
/codon\_start=1  
/product="putative regulatory protein"  
/translation="MRDEPTARRRRLDGGESRRRQQAEGEGNDGATRGFLAV  
RSTRFELDASGLDGTSEAGDVLVRSDDGEGHTASDGNRRGNSLSGLRP  
KAKTERAPKPSASSNAPVLPALEGQRDRARGRQHWGGEWKEKGGKRLGEDDA  
GEERKGEKCKGLPICRFGKKEEVSDEAEGGMLPESLGLREAWGRVNTTAM  
TAGRFGARRRRRAAGDGGRAAGDRVAQTRDRTGGNHAGARANERGER"  
complement (<16478, .>23409)  
/locus\_tag="OSJNBa0073D04.35"  
/note="similar to gag-pol protein GB:AAL59229 GI:18092337  
(Zea mays)"  
complement [join(<16478, .17166,17233, .21196,21628, .22437,  
23383, .>23409)]  
/locus\_tag="OSJNBa0073D04.35"  
complement [join(16478, .17166,17233, .21196,21628, .22437,  
23383, .>23409)]  
/locus\_tag="OSJNBa0073D04.35"  
/codon\_start=1  
/product="putative gag-pol protein"  
/translation="MPTVMGCEMAYNANFSSHHGFGHHVQRLDLARFLGSPCEY  
TGYSDVPPHCELSVLYPRGGIHGAGLPHHFIVARPTLQIAYQDLSWTALRLAH  
DYSRLGGSAPQPLRPSGSHDARPCYSEOPVLVWVLSAQATQDHLILKAY  
RLARRVAALEARVLAEGVTSVDITSWDSGLCHASHLSHSSRGSAATPSGPRSPSS  
RAPYSPVSPVPHVPSYTPGHPVASSRGCFIETARKVYGVSTVREFHLAPP  
PPAMASPRRAACAPARFGEGRVKEANESHQSNRSHRIVSNNAKVEQPC  
RSNKRSHRTASDMEQSHSRHTASHNSGEECLPSPHVNNQHLMVQTOQLGI  
AAVAGYKNSHSGHPMGNRSKLTDFLRSPPEFSQTPVEPADDDMLKDYDRKIN  
LVQCTPQKTLIYASHQIRGPAANWENYCNAPHEPTNIADFEATPAAHVPESITD  
MKCKEFLRLKQNSVNEYL.SMFNKLARYALEVDSDDKKIRKFLKGIAGVGLQLLA  
HDPTFOHMKALLLEDARKEATEYKESKSHQNSRGAAPRYPGQPMOYHOSVT  
QANRPGVAPRPPQOQHAPSGNTAPNSFTSKSPQSPVAVQCFRCQMGHY  
ARQCPQNTNSGHANGSARTPTPAAASRUSTQASGGSASNSHFGRCVGNHLRA  
ETAGADPVVNGSVNSVPAVLFDGSGHSPISQAFVKNRKNQTNLRVPMVHSP  
GNIRATQICEVNLRIEVDFLAKPVLDSQSLDILGDMWLAKHQDCKAEKST  
LQCGPGQVRTPPTPASRLTCLQVTSLESPVICEYDPVPELTPMPDRRIE  
FAIELAPCTAIKRPQMAANELTEVKRQIGESKGVYRPPSPMGAPVLLVKKKD  
GSEBWDYDALNEVTIKNKYPLRIDDLFDQLGARVFSKIDLRSGYHOLKIRSDI  
PKTASTYGLYETVMSFGLTNAPTFMNLKIFMEYLDQVWFIDILLIYSKNE  
EBAEHLRLNEKLRDQLFAKSKCFWLDRAVFLGHVISSNGVEVDPKSVAVLAW  
NPPKNVSEISFGLAGYRRFIFGSLKARLPMTELLKKEKFRWSAACDSFOEKK  
RLTTAPVLTUPDIRKDFEIFCDASRQGLGCVLIQERKVAYASRQLRPEVNTPTDL

ELAAVVHAKIMREHYLIENRCEVYTDHKSILKVIPTQTELNMQRWLEILKDYDILGIH  
YIPGKANVVDALSRLKTYCNVQIRPDQDHLCKRELEKRLTIVQSGVPASLTIVQPTILE  
SQIREAQDDEGKELIKRIQEKQNTFISDQGTWCGPRICVFAKKEQRAGLILKEA  
HESAYSIHFGSTKMTQDIKAYFWAGMKEDVAYVALCDICQVRAEAKQRAGLILKEA  
PIPEWKEEIGMDFITGLPRTPNGYDSIWIIVDRITKSAHFVPTVTTFGKGLALYM  
THVCRFGCPKKIVSDRQFQTSRFEWKOLHEALGTDLNFMAHYPDCTGTDWRCPLPYA  
EFSYNSYQASITHMSNEAMFGKCHTPLCMNEVEALVFGPDILKSAEEOVKLIRER  
LKTANRQKSYADNRDRDLFEKGDHVLVRSPLSGMRFGMSGLAPRYIGPYLITA  
RGEVAYQUELEPDLADVNFHVSQLKKLRVPEEQVPLGNIVLEKNLYKERFIKV  
LBEAEKQTRTKIKFYKVQVNSHSEDEATWEREDLLRAEFPBLLP"  
/locus\_tag="OSJNBa0073D04.39"  
/note="predicted by genemarker"  
join(<25617, .25780,25922, .>26162)  
/locus\_tag="OSJNBa0073D04.39"  
join(25617, .25780,25922, .>26162)  
/locus\_tag="OSJNBa0073D04.39"  
/codon\_start=1  
/product="hypothetical protein"  
/translation="MSVNEPTCAVKGASRVAKAATGRMTTPARKKKGKRGKGA  
CPFAQSGKRRRVQRLTAAGTSATTARGLQREARMEWRGSGTRAGSGQAAAAA  
LKAAGORDRAAQTRDRTGGNHAGARANERGER"  
/locus\_tag="OSJNBa0073D04.37"  
/note="similar to nuclear speckle type protein SP0P  
GB:CAA04199 GI:2695708 (Homo sapiens)"  
join(<27867, .28919,29563, .29682,30319, .30366,30460, .30580,  
31647, .31703,32683, .32791,32872, .33097,33182, .33298,  
33615, .33693,33877, .33899,34272, .34316,34446, .>34604)  
/locus\_tag="OSJNBa0073D04.37"  
join(27867, .28919,29563, .29682,30319, .30366,30460, .30580,  
31647, .31703,32683, .32791,32872, .33097,33182, .33298,  
33615, .33693,33877, .33899,34272, .34316,34446, .>34604)  
/locus\_tag="OSJNBa0073D04.37"  
/codon\_start=1  
/product="putative speckle-type protein"  
/translation="MSSPSSQOPASAIKAPTTSYGYRLRIDYVRSIGSPGWSLRSD  
FVVGSGRQFISYVNGNRPENAEFISVLCDDSSPKPAMLOVITFDDEAKKQSLR  
KAPVTIATACAGCHRVKRDRLAKSKIRDFGFTTICDVLIDHFTAQDEDFVY  
SVPPERDLGLDITSGGDVVFQVGGAEFAHRLGLLAASPVLAALYIGPMWEGG  
GLQGGVAKIDDDMLVFLKRLRYATPSLPQMQQGLEEGRAVAQHLLAAADRYG  
MERLRLICQAQCKHIEVASVASILLIADQHGCSGLKNAKCFELKSPGFAAAMATQE

Query Match 71.9%; Score 19.4; DB 8; Length 300029;  
Best Local Similarity 95.2%; Pred. No. 1.5e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 GCCTTTAGATGACACCTG 26  
|||  
Db 56593 GCACCTTAGATGACACCTG 56613

RESULT 6  
EX005194/c  
LOCUS 151076 bp DNA linear VRT 23-SEP-2003  
DEFINITION Zebrafish DNA sequence from clone CH211-188A7 in linkage group 25,  
complete sequence.  
ACCESSION EX005194  
VERSION EX005194.10 GI:34996469  
KEYWORDS HTG.  
SOURCE Danio rerio (zebrafish)  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 151076)  
AUTHORS Giselte H.  
TITLE Direct Submission  
JOURNAL Submitted (23-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,





```
OY 2 ACCCGCGCTTTAGGATCGACACCTG 26
|||||
Db 783 ACCCGCGCTTTAGCACCGACCTG 807
|||||

RESULT 9
AF548445/c
LOCUS      2851 bp      DNA      linear      BCT 08-APR-2003
DEFINITION Uncultured bacterium plasmid pAK206 unknown genes.
ACCESSION  AF548445
VERSION     AF548445.1 GI:29611371
KEYWORDS   uncultured bacterium
SOURCE     uncultured bacterium
ORGANISM   Bacteria; environmental samples.
REFERENCE  1 (bases 1 to 2851)
AUTHORS   Knetsch,A., Waschowitz,T., Bowien,S., Henne,A. and Daniel,R.
TITLE     Metagenomes of Complex Microbial Consortia Derived from Different
          Soils as Sources for Novel Genes Confering Formation of Carbonyls
          from Short-Chain Polyols on Escherichia coli
JOURNAL   J. Mol. Microbiol. Biotechnol. 5 (1), 46-56 (2003)
MEDLINE   22560762
PUBMED    12673061
REFERENCE  2 (bases 1 to 2851)
AUTHORS   Knetsch,A. and Daniel,R.
TITLE     Direct Submission
JOURNAL   Submitted (26-SEP-2002) Institut fur Mikrobiologie und Genetik,
          University of Goettingen, Grisebachstr. 8, Goettingen 37077,
          Germany
FEATURES   Location/Qualifiers
            source
              1. .2851
                /organism="uncultured bacterium"
                /mol_type="genomic DNA"
                /isolation_source="soil"
                /db_xref="taxon:77133"
                /environmental_sample
                /lab_host="Escherichia coli"
                /plasmid="pAK206"
                /note="anonymous microbial sequence conferring formation
                of carbonyls from short-chain polyols on Escherichia coli"
              94. .1092
                /note="orf10"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAO91882.1"
                /db_xref="GI:29611372"
                /translation="MCMPGGTLISSPWSCTRRSGRKGKISPKTLASAEKDATFTSGG
                TSVRLWRTSPVPSLGGSTDDTSVATSTWLSNRPSTASSRRCLTAGASIGSDRMT
                MRLPATSAACVTRERLGTTPSHATSLPWGGPOPAHLGTWSPVSVRYIARSRG
                RAGLRHPRPSAVLQREGPASAGPRRAGSPPLPQKFSQNVAAAWTASRISSRR
                LPFAGVLISNGAYSCAVIKSRPAVSRSPLMLTAVPYGKNRVIGVWGHAPPAIPPI
                TNSTIVRPNTAAMARLRMPYDVCAAVSARASSARSARFVSEAAFSNSVRASA"
                complement(984. .2171)
                /note="orf11"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAO91883.1"
                /db_xref="GI:29611373"
                /translation="MVRDVGAEADLAQALVAALQEPWASGIPRNPAGMLTAAAKRRA
                IDHFRSLDKRKEELGVAIEBTGOERAPDADDDVDGDLRLIFTACHPVLSIT
                EARLALRLGLLTETIEIAFLVAEPTIAQIRVAKRTIAEKRVFVPRGSEVE
                ELASLVLYLIFNGEYATAGDGLIRSLCEDALRLGLVLAGLAPDEPEVHGLVLM
                EIOAKTRARVGTGEPVVLDDONRALMDRLRLLGRGLAALQALGAGFGYILQVA
                IAACHARALEPQRTWRIVALTTELLHTRSPVVDLNRVAVAFAGPAAGFYLVQAA
                LRLEPALAEYHLLPSVRGDLAKGRHAEARTEFERAASLTNRNAREALLARAETRA
                QTS"
                complement(2519. .2707)
                /note="orf12"
                /codon_start=1
                /transl_table=11

ORIGIN
Query Match      68.9%; Score 18.6; DB 1; Length 2851;
Best Local Similarity 84.0%; Pred. No. 4.le+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 3 CCCGCGCTTTAGGATCGACACCTGA 27
|||||
Db 1445 CGCGCGCTTTGGGATCGCTCTCTGA 1421
|||||

RESULT 10
AE014432
LOCUS      11481 bp      DNA      linear      BCT 30-SEP-2002
DEFINITION Brucella suis 1330 chromosome I section 118 of 190 of the complete
          genome.
ACCESSION  AE014432
VERSION     AE014432.1 GI:23348184
KEYWORDS   Brucella suis 1330
SOURCE     Brucella suis 1330
ORGANISM   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
          Brucellaceae; Brucella.
REFERENCE  1 (bases 1 to 11481)
AUTHORS   Paulsen,I., Seshadri,R., Nelson,K.E., Eisen,J.A., Heidelberg,J.F.,
          Read,T.D., Dodson,R.J., Umayam,L.A., Brinkac,L.M., Beanan,M.J.,
          Daugherty,S.C., Deboy,R., Durkin,A.S., Kolonay,J.F., Madupu,R.,
          Nelson,W.C., Ayodeji,B., Kraul,M., Shetty,J., Malek,J.A., Van
          Aken,S.E., Riedmuller,S., Tettelin,H., Gill,S., White,O.,
          Salzberg,S.L., Hoover,D.L., Lindler,L., Halling,S.M., Boyle,S.M.
          and Fraser,C.M.
TITLE     The Brucella suis genome reveals fundamental similarities between
          animal and plant pathogens and symbionts
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99, 13148-13153 (2002)
REFERENCE  2 (bases 1 to 11481)
AUTHORS   Paulsen,I., Seshadri,R., Nelson,K.E., Eisen,J.A., Heidelberg,J.F.,
          Read,T.D., Dodson,R.J., Umayam,L.A., Brinkac,L.M., Beanan,M.J.,
          Daugherty,S.C., Deboy,R., Durkin,A.S., Kolonay,J.F., Madupu,R.,
          Nelson,W.C., Ayodeji,B., Kraul,M., Shetty,J., Malek,J.A., Van
          Aken,S.E., Riedmuller,S., Tettelin,H., Gill,S., White,O.,
          Salzberg,S.L., Hoover,D.L., Lindler,L., Halling,S.M., Boyle,S.M.
          and Fraser,C.M.
TITLE     Direct Submission
JOURNAL   Submitted (14-AUG-2002) The Institute for Genomic Research, 9712
          Medical Center Dr, Rockville, MD 20850, USA
FEATURES   Location/Qualifiers
            source
              1. .11481
                /organism="Brucella suis 1330"
                /mol_type="genomic DNA"
                /strain="1330"
                /db_xref="taxon:204722"
                /chromosome="I"
                /locus_tag="BR1343"
                /locus_tag="BR1343"
                /locus_tag="BR1343"
                /note="similar to GP:15073285; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="oxidoreductase, putative"
                /protein_id="AAO30257.1"
                /db_xref="GI:23348185"
                /translation="MAVIAGLIGSGFMGKTHALANVRIFDPLAVOLHUTGDDIV
                QTKAAROLGFAHATGDNRDLISNPDINLITTPNRFKEMALAAIKAGKHVYCEK
                PLAPSAADCLNTLAEREGIOIOTAVGNVKNPKMMLAOEIIASGELGTRNPRGTHA
                EDEWADARIPWTWRIDPAGGGALADIGSHIATARYLIGTIADVNGEAVTLIREPD
                MADLSRTKPEVDVDCVAFMRFENGATGILEASWASVGRMQHDFEISGSGKAIFFSQ

/product="unknown"
/protein_id="AAO91884.1"
/db_xref="GI:29611374"
/translation="MMIVSKDKKSEAGILLPDVENAGRHGQVQRNNGEGRDAGWGRTA
GELEGRARPVLQGPFRQ"
```

```

ERNLELELYSDRGQKQFRKI FAPGPEHPYGATCVAGGHOIGFNDLKTIEVRDFVL
AIAGQTKHADFRGVEYQKVTETIYASSRQWKKI"
complement (1332..2108)
/locus_tag="BR1344"
CDS
complement (1332..2108)
/locus_tag="BR1344"
/notes="identified by match to TIGR protein family HMM
TIGR01187"
/codon_start=1
/transl_table=1
/product="iron compound ABC transporter, ATP-binding
protein"
/protein_id="AAN30258.1"
/db_xref="GI:23348186"
/translation="MTLLSVKMLDVLGKRALDASFETKGGFGLGVNGAGKTT
LIRAIAGLVASSQVSLAGRDLEHMGAEKARSLAYLPOERDVAVFVHMLVLSGRS
ALKPFPAGLDQDEAIIEIAMDMDVSRFGERSVMELSGGERARVLIARVLAQDTPVI
LADPEVAGLDPAHQLTMEITFAELAREGRTVTIASLHLSLAACHSRLILLDHGCMAA
DGPFAVFTPQRLRDVYGRARIMVMDGSEFVHPTALQPR"
complement (2105..3106)
/locus_tag="BR1345"
CDS
complement (2105..3106)
/locus_tag="BR1345"
/notes="identified by match to PFAM protein family HMM
PF03772"
/codon_start=1
/transl_table=1
/product="iron compound ABC transporter, permease protein"
/protein_id="AAN30259.1"
/db_xref="GI:23348187"
/translation="MTERHFHLLAGLALLVILFAVSLTGPAAIGIGSFKALLG
DMDTVLWMRIIRPALLALLIGASGLGAOLGYLRNPLAEPGLGVASASLG
AVTAIYSGLSLFPALPLALAGAFVSVELAKLAGRNAGTAVILAGVATVSLAGA
LITALNLSPNPANIEIVFMWLSLADSNTHVALVPPILLGLMELLSLGRSDLSL
TIGSDAAATGWSLGRVQVFAVLGTRACVAGSTAVAGSIGFVGLVPHLLRVLGARP
SRLVASGIGGAAALLDAADILRVVMPGRELKGLVITALLIGAPFLMLVFKTRQLV"
complement (3099..3938)
/locus_tag="BR1346"
CDS
complement (3099..3938)
/locus_tag="BR1346"
/notes="identified by match to PFAM protein family HMM
PF01497"
/codon_start=1
/transl_table=1
/product="iron compound ABC transporter, periplasmic iron
compound-binding protein"
/protein_id="AAN30260.1"
/db_xref="GI:23348188"
/translation="MKRVSVKVTBRLCLAAGLALFAPLPANAADIAQBVVSVNCTDQ
LAVLAREQOLSVSLSDPDQLSVNADKAERLPINHAQAEVFLQKDDVLVAGTFSS
RATVGLRLGVIRVEFAPSFEDIQEHRLRIGELGQRAQAQIDDMKAALAIK
RPHAKRVATYANSVSGQTLVDVFAVLGNAGNAGEIGITGSALLPLEKLVLEKP
DILVRSRRAPALAPENFQHPALRALEKQARAI SMADNLTVCGGPPFVFAELAEAE
ARD"
complement (3952..5814)
/locus_tag="BR1347"
CDS
complement (3952..5814)
/locus_tag="BR1347"
/notes="identified by match to PFAM protein family HMM
PF04052"
/codon_start=1
/transl_table=1
/product="iron compound TonB-dependent receptor, putative"
/protein_id="AAN30261.1"
/db_xref="GI:23348189"
/translation="MHILKSALLASACATLSPALAQDGKDDGVLTDTIVTPPLR
RASSLQRTSSVSDIADBIERSAADPLQSLQTVSGVSKTNGQSGSSANIMYRGM
SKOTVVLVNGVRTASGTALANPLTSIERIARIAGHASSQYGAIDAIGGVNIIT
KQSGACGERACWCSVGTGVSHKPGVSGSLQGRSSDGDYAVGAFTGTQGYDTTP
EAFCHPDDQFLQGSFNALSKDPWKGIVADGLFSRGNQYDAPAFNEADSTAF
TKGVGTRIDHTADWSVTFSGIDNSRNFKIGESDRFETRRYGVFASTEKSPDTG
KVSHVTVGGVAREKINTIIDYETGRDLAAVFGQISLEYDALRFDGRIYDHNGQF
GNVTTYNLGASVEILPDILRSYATGFRAPTTNLYEPGFANPDLOPEKSRSVGL

```

```

NMQATASLDMALYQTRLSDAITSPSYIPYNTASAKVTGLEATLSHRENEOMGKIK
GMWDLKRPVDEDSNDLPYRERFKAAAEVNVFPEKFLDLTARVLGGSGYIYAKNTKK
LGDYVTADFVALYSIDKQSLFSEVENIFDKOYETSSGVAFGRITIIIGLTRNF"
complement (6103..6258)
/locus_tag="BR1348"
CDS
complement (6103..6258)
/locus_tag="BR1348"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAN30262.1"
/db_xref="GI:23348190"
/translation="MSSRSIARFGAVANFQSVCYRNFQETVPPLRMKREHGEATLKG
PLRPPQL"
complement (6308..7420)
/locus_tag="BR1349"
CDS
complement (6308..7420)
/locus_tag="BR1349"
/notes="similar to GP:15158527; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="ABC transporter, permease protein, putative"
/protein_id="AAN30263.1"
/db_xref="GI:23348191"
/translation="MRIANFRIGIKELRGLMRDPMLLIIVVAFTLAIYTAAKAMPE
TANAALAIVDEQDSQSVGRITAFYVPVFPKLISSYEMDNMDGDLTALNIPP
DFQDILLAGRAPTIQLNATMSOAFGTGGVQSVIAGEVEFLNHYGSGAPAVDL
NLSARINKWNGFCAINIVISSYTNLSIILTGALIRERHGHIEHLIWMVPTLE
IMSKWSMGLVWVASAFALTFFVQGLSVINGSILLFLAGAAQLFATLSGLFPL
AIVAGNPGQGLLMLLIPLOVLSGATTPRESMPEIIQITMLAENPHTFVILAOSIL
FRAGAGCVWMPQFLMFLVIGSVLFLFALRRFRAFLR"
complement (7422..10190)
/locus_tag="BR1350"
CDS
complement (7422..10190)
/locus_tag="BR1350"
/notes="similar to GP:15158528; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="ABC transporter, ATP binding/permease protein"
/protein_id="AAN30264.1"
/db_xref="GI:23348192"
/translation="MSSVFGKAGDGGFVRARVDLSYGATRALRNVDLDVAGMVG
LIGPDGVGSSLSLIAGATIQQGHVGLGGDIADKRRQSAIPRIAYNPQGLGKQL
YPTLSVENIDFFGRFGHDRAERERRIADLLARTGSAFANRPAKLSGGMKQKTSL
CCSLIHDPDLVIDEPTGVDPLSRQFWELIIDIRKDRPGMSVIVATYAMEAEERFD
WLVAMDGOILATGTQELLERTDTNLDAAFTALLPEQRGRHREIHIPRAPASDT
AIALEAHVTVRFSDTAVDNVSFRI PRGEI FGLGNGCGKTTTKVLTGLLPASEG
TALFREGVDPKIDIRRVAGVMSQAFSLYSELTVQNLEHARLFGMDPDTVAREIK
ELTARFDLADMDLEDPMLGTRQRLSLAVALIHSPDILIDEPSTGSDVPVARDAPW
QIILADSRDNVTIFVSTHFMNABECLDRISMHAGKVLISDTPKAITSRNAATLED
AFVAFQDAI QEGEQAPQRIEKDSAEPAQRKAWPSLNLNRRFATRRAELELK
RPIRATLALIGSVILMFGVIGLNDVEKLTFAVLRDQDTTISREYTOIAGSRFYI
EXPITDYDOLDRMRGEISLAIETPPNFAANVMRGQVEVAAMIDGAMPRAETAK
GVYQGVHANWLTOKARALYGDATAGNENLEIRYVNPVQSVLAVMVAIVILLIMI
PAMVLVSVREXELGSLINFYVTPVTKFEFLGKQLPVIGLAFINFLMTAPVFIIF
RVPTFSFVYTAALLYIIITSMGLVISTFNMNQIAAIFGTALMTLIPAVQYSGLI
DPVSALQGVGAFIGKIYPTTYFTVISRGTFSKALDFSALAGSFLPLLIILAIPLMLLGI
VLKKQAS"
complement (10187..11254)
/locus_tag="BR1351"
CDS
complement (10187..11254)
/locus_tag="BR1351"
/notes="similar to GP:15158529; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="HlyD family secretion protein"
/protein_id="AAN30265.1"
/db_xref="GI:23348193"

```

	Query Match	68.9%; Score 18.6; DB 1; Length 11481;
	Best Local Similarity	84.0%; Pred. No. 4e-02;
	Matches	21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy	3 CCGCGGCTTTAGGATCGACACTGA 27       	
Db	8852 CCGCGGCCGGATCGACACTGA 8876       	
RESULT 11		
AE009507/c		
LOCUS	13916 bp DNA linear BCT 20-MAR-2003	
DEFINITION	Bruceella melitensis 16M chromosome I, section 64 of 195 of the complete sequence.	
ACCESSION	AE009507 AE008917	
VERSION	AE009507.1 GI:17982577	
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		
1 DelVecchio,V.G., Kapratir,V., Redkar,R.J., Patra,G., Mujer,C., Los,T., Ivanova,N., Anderson,I., Bhattacharya,A., Lykidis,A., Reznik,G., Jablonski,L., Larsen,N., D'Souza,M., Bernal,A., Mazur,M., Goltsman,E., Selkov,E., Elzer,P.H., Hagius,S., O'Callaghan,D., Letesson,J.-J., Haselkorn,R., Kyripides,N. and Overbeek,R.		
TITLE	The genome sequence of the facultative intracellular pathogen Bruceella melitensis	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)	
PUBMED	11756688	
REFERENCE	2 (bases 1 to 13916)	
AUTHORS	DelVecchio,V.G., Redkar,R.J., Patra,G. and Mujer,C.	
TITLE	Direct Submission	
JOURNAL	Submitted (13-NOV-2001) Institute of Molecular Biology and Medicine, University of Scranton, Scranton, PA 18510, USA	
REFERENCE	3 (bases 1 to 13916)	
AUTHORS	Elzer,P.H. and Hagius,S.	
TITLE	Direct Submission	
JOURNAL	Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag Center, 111 Dalrymple Building, Baton Rouge, LA 70803, USA	
REFERENCE	4 (bases 1 to 13916)	
AUTHORS	Kapratir,V., Los,T., Ivanova,N., Anderson,I., Bhattacharya,A., Lykidis,A., Reznik,G., Jablonski,L., Larsen,N., D'Souza,M., Bernal,A., Mazur,M., Goltsman,E., Selkov,E., Haselkorn,R., Kyripides,N. and Overbeek,R.	
TITLE	Direct Submission	
JOURNAL	Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell Park Drive, IL 60612, USA	
REFERENCE	5 (bases 1 to 13916)	
AUTHORS	Letesson,J.-J.	
TITLE	Direct Submission	
JOURNAL	Submitted (13-NOV-2001) Unite de Recherche en Biologie Molculaire, Laboratoire d'Immunologie et de Microbiologie, Universite of Namur, 61 rue de Bruxelles, Namur 5000, Belgium	
REFERENCE	6 (bases 1 to 13916)	
AUTHORS	O'Callaghan,D.	
TITLE	Direct Submission	
JOURNAL	Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue Kennedy, Nimes 30900, France	
FEATURES		
source		
1..13916	/organism="Bruceella melitensis 16M"	
	/mol_type="genomic DNA"	
	/strain="16M"	
	/db_xref="taxon:224914"	
	/chromosome="I"	
	complement(78)..458)	
gene	/gene="BMEI0650"	
CDS	complement(78)..458) /gene="BMEI0650" /codon_start=1	

/gene="BMEI0655"  
/codon\_start=1  
/transl\_table=11  
/product="ABC TRANSPORTER ATP-BINDING PROTEIN"  
/protein\_id="AAL51836.1"  
/db\_xref="GI:17982583"  
/translation="MTISTCECAKASPLKYRIRISOLMSGGQROVEVAWIDGAMP  
TRAETVGVQVGHANLWTKARLYGDAATAGNENLIRYRNPVDVOSLVAMYPAVI  
PILLMLAVMEKVLVSVKEKLSIINFVPTVKFEELGKOLPYICLAFNLMLT  
AFVFTFRPFTGCVTVATALLVYITTSKGLVSTFMSQIAAIEGTALMTLPA  
VOYGLIDVSALQGVAFIGKIPITTYITTSKGLVSTFMSQIAAIEGTALMTLPA  
VIMLLGIVLLKQAS"  
6475..7587  
/gene="BMEI0656"  
6475..7587  
/gene="BMEI0656"  
/codon\_start=1  
/transl\_table=11  
/product="DAUNORUBICIN RESISTANCE TRANSMEMBRANE PROTEIN"  
/protein\_id="AAL51837.1"  
/db\_xref="GI:17982584"  
/translation="MRIANIPIRIGIKELRGLMRDPMMLLILIVYATFLAIYTAAKAMPE  
TLNNAIAIVDEQSPVSRITTAEPVFPVKLISVYEMDNRMDEGLDTFALNIPP  
DFORDLAGRATQINVDATMSQAFTGGVYQSIIVAGEVSEPLNHYEGSAPAPVOL  
NURARFQNLNGWFGAITNVISSVTMLSLITLGAALRERHGTIEHLNVPYPIE  
IMISKVMSGLVIVASAFALTVVQGLSVPIGSIILFLAGAALQLPATISUGIFL  
ATVAGMPOGLMLLILPLQLVLSGATTPRESMEPIITQINLAAPNTHFVILAQSIL  
FRGAGLVVWPOFLALFVIGSVLFLFALRRFAFLR"  
8144..9943  
/gene="BMEI0657"  
8144..9943  
/gene="BMEI0657"  
/codon\_start=1  
/transl\_table=11  
/product="METAL CHELATE OUTER MEMBRANE RECEPTOR"  
/protein\_id="AAL51838.1"  
/db\_xref="GI:17982585"  
/translation="MAQDGGDDGVDLTIVTILRASSLQRTSSVSVIDRADIE  
RSAAPDLQQLQYSGVKTNGQSSADIWRGSSKQIVLVNVRTASATSGST  
ALANIPLTIERIETARGISQYGAADITGVINIIITQGGACGERACGVSITGVSH  
PMGYASGLQSSGIDYAVGAFTQGTQDFTTPEAFGEHPDDGFLQGSFNFAL  
SKDFMGKIVKDFGRDNQYATAPAFNEADSTAFKGTVIDHTADWSSTVEFS  
TGIDNSRNPKEGSDWETRYGVFASTKSFDTGVSHVTVGGVAYREKINTII  
DYDETGRDLAIVPGQSLSDYDALRFDGGIRYDHNGQFNVTYNGASVEILLPLDLR  
SVATGFRATNELIYFGAPNDLQPEKRSVEVGLNMQATASLSMALYQTELSL  
AINSTAPYIPYNAKVTGLEATLSHFNQWGIKGMVDLRVDESDSGNDLYRE  
RFKAAAEVNFKEKDLTLARVLYGGSRYTNNAKTKLGDYVTDADFVALYSIDKOSQL  
KFSVENIFDKDYETSSGVYAPQRTITIGLTRNF"  
9957..10796  
/gene="BMEI0658"  
9957..10796  
/gene="BMEI0658"  
/codon\_start=1  
/transl\_table=11  
/product="METAL CHELATE PERIPLASMIC BINDING PROTEIN"  
/protein\_id="AAL51839.1"  
/db\_xref="GI:17982586"  
/translation="MKRVSVKWTRELCIAGLALFAPLPANAADIQAQVWSINVCTDQ  
LAMLAREGQLQSVSLSDPQLSVWADKAEPLPINHAQAEVFLQKPLDLVLAGTFSS  
RATVGLLRIGIRVEEFAPRSDFQEHRLRIGELLGRLQAEQAQIQIDDMKALAAIK  
RPDHAKVYALYANYSVQGGTLVDEAVLAGLENWAGNAGEIGITGSGALLPLEKLVLEKP  
DILVRSRDRAPALAFENFQHPALRALEKRAISNADNLTVCGGPFSSVEVAELAEAE  
ARD"  
10774..11790  
/gene="BMEI0659"  
10774..11790  
/gene="BMEI0659"  
68.9%; Score 18.6; DB 1; Length 13916;  
Best Local Similarity 84.0%; Pred. No. 4e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
3 CCGCGCGTTTAGGATCGACACCTGA 27  
|||||  
5042 CCGCGCGCAGGGATCGACACCTGA 5018  
|||||

RESULT 12  
AB007645 72698 bp DNA linear PLN 27-DEC-2000  
LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K8K14.  
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K8K14.  
ACCESSION AB007645 BA000015  
VERSION AB007645.1 GI:2564045  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
1 (sites)  
Nakamura,Y., Sato,S., Kaneko,T., Kotani,H., Asamizu,E., Miyajima,N.  
and Tabata,S.  
Structural analysis of Arabidopsis thaliana chromosome 5. III.  
Sequence features of the regions of 1,191,918 bp covered by  
seventeen physically assigned pl clones  
DNA Res. 4 (6), 401-414 (1997)  
98162728  
2 (bases 1 to 72698)  
Nakamura,Y.  
Direct Submission  
Submitted (03-OCT-1997) Yasukazu Nakamura, Kazusa DNA Research  
Institute, Department of Plant Gene Research, 1332-3, Yana,  
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,  
Tel:81-438-52-3935, Fax:81-438-52-3934)  
Address for correspondence: kaos@kazusa.or.jp  
For the latest information on annotation of this clone, please see  
http://www.kazusa.or.jp/kaos/cgi-bin/sgd/graph.cgi?c=K8K14  
Genes with similarity to proteins in the databases are described in  
'product' or 'note' qualifiers. Genes that have no significant  
protein similarity are described as 'unknown protein'.  
The software programs used to predict genes include: Grail  
(Informatics Group, Oak Ridge National Laboratory,  
http://compbio.ornl.gov/Grail-1.3/),  
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),  
Negen2 (S.M. Hebsgaard, et al., CBS, Technical University of  
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and  
Splice predictor (Volker Brendel, Stanford University,  
http://grm1.ni.zoel.iastate.edu/cgi-bin/sp.cgi).  
Genes encoding tRNAs are predicted by tRNAscan-SE  
(Sean Eddy, Washington University School of Medicine, St. Louis,  
http://genome.wustl.edu/eddy/tRNAscan-SE/).  
This sequence may not be the entire insert of this clone. It may be  
shorter because we remove overlaps between neighboring submissions.  
The 5' clone is K3G17 and the 3' clone is K919.  
Location/Qualifiers  
1..72698  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="K8K14"  
/clone\_lib="Mitsui TAC"  
complement(join(283..653,770..1619))  
/note="contains similarity to unknown protein  
gene\_id:K8K14.1  
sp|Q10058"  
/codon\_start=1  
/evidence=not experimental  
/protein\_id="BAB09014.1"  
/db\_xref="GI:9758428"  
/translation="MAVISLPPVLTTRITTCNSKSKMAKQIGETSRRVVF  
CGGVIGCTATYFLAKKGIATLVKESAVACAGGAGGFLAFDWCDSVVAHLARAS  
FLHRLAEELNGVESYGYRALTIVSTVESKGGGLGLPWNVNGPAKSTIGTT  
QTAAQVHPQLFTRKLLSTAREKYGIVICKEVRVESRVNSVLEGGRVLDVW  
VLAMPWSKSLSSIFRYVCTKADHSIVLEKPEPNATTHALFLTRPAGGALDP  
EYVPRPTGEVYICGSSQEEVDDADQVTSNQVLTQVLRVAKTVTSYLNENALVKA





6 (bases 1 to 162646)  
Worley, K.C.  
Direct Submission  
Submitted (25-JAN-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Mar 1, 2002 this sequence version replaced gi:17977513.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
features listing.

ANNOTATION OF FEATURES:  
STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.  
Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.  
Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL:  
<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES	source
misc_feature	Location/Qualifiers 1..162646 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="3" /clone="RP11-600G3" 1..2004 /note="overlaps bases 79511..81510 of clone AC117486" /function="clone overlap"
repeat_region	complement(248..339) /rpt_family="MIR"
repeat_region	491..556 /rpt_family="(TA)n" complement(844..981) /rpt_family="THE1B"
repeat_region	1586..1727 /rpt_family="MIR"
repeat_region	1683..1736 /rpt_family="L2"
repeat_region	2734..2922 /rpt_family="MIR"
repeat_region	2977..3083 /rpt_family="L2"
repeat_region	complement(3383..3562) /rpt_family="MIR"
repeat_region	3742..7012 /rpt_family="L1MC3"

```

TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 2128 20-JUN-2001;
          KIOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
  source
    1. .1287
      /location/Qualifiers
      /organism="Corynebacterium glutamicum"
      /mol_type="unassigned DNA"
      /db_xref="taxon:1718"
ORIGIN
  Query Match 66.7%; Score 18; DB 6; Length 1287;
  Best Local Similarity 80.8%; Pred. No. 8.4e+02;
  Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 CACCCGCGCTTTAGGATCGACACCTG 26
    |||||
Db 90 CAGGCGCTCTTTAGGAGCGACACTG 65

RESULT 16
BD164329/c
LOCUS BD164329 Novel polynucleotide. 1287 bp. DNA linear PAT 17-JAN-2003
DEFINITION Novel polynucleotide.
ACCESSION BD164329
VERSION BD164329.1 GI:27870141
KEYWORDS JP 2002191370-A/2128.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 1287)
AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoo, A., Ikeda, M. and Ozaki, A.
TITLE Novel polynucleotide
JOURNAL Patent: JP 2002191370-A 2128 09-JUL-2002;
          KYOWA HAKKO KOGYO CO LTD
COMMENT OS Corynebacterium glutamicum
PN JP 2002191370-A/2128
PD 09-JUL-2002
PF 15-DEC-2000 JP 2000405096
PI SATOSHI NAKAGAWA, HIROSHI MIZOGUCHI, SEIKO ANDO, MIKIO HAYASHI,
PI KEIKO OCHIAI,
PI HARUHIKO YOKOI, NAOKO TATEISHI, AKIHIRO SENOO, MASATO IKEDA, AKIO
PI OZAKI
PC C12N15/09, C12N15/09, C07K14/34, C07K16/12, C07K16/40, C12M1/00, PC
C12N1/15,
PC C12N1/19, C12N1/21, C12N5/10, C12N9/00, C12N9/02, C12P7/40, C12P13/
PC 04, C12P13/08,
PC C12P19/00, C12P19/34, C12P21/02, C12Q1/37, C12Q1/68, G01N33/53, PC
G01N33/566,
PC G01N33/569, G01N33/68, G01N37/00//C12P21/08, (C12N1/21, C12R1:15),
PC (C12N1/21, C12R1:13), (C12N1/21, C12R1:01), (C12P13/08, C12R1:15),
PC C12N15/00,
PC C12N5/00, C12N15/00
CC Novel polynucleotide
FH Key
FT source
FT Location/Qualifiers
  source
    1. .1287
      /location/Qualifiers
      /organism="unidentified"
      /mol_type="genomic DNA"
      /db_xref="taxon:32644"
ORIGIN
  Query Match 66.7%; Score 18; DB 6; Length 1287;
  Best Local Similarity 80.8%; Pred. No. 8.4e+02;
  Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 CACCCGCGCTTTAGGATCGACACCTG 26
    |||||
Db 90 CAGGCGCTCTTTAGGAGCGACACTG 65

RESULT 17
AX067039/c
LOCUS AX067039 1335 bp. DNA linear PAT 24-JAN-2001
DEFINITION Sequence 621 from Patent WO0100805.
ACCESSION AX067039
VERSION AX067039.1 GI:12544747
KEYWORDS Corynebacterium glutamicum
SOURCE Corynebacterium glutamicum
ORGANISM Corynebacterium glutamicum
REFERENCE 1
AUTHORS Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.
TITLE Corynebacterium glutamicum genes encoding proteins involved in
membrane synthesis and membrane transport
JOURNAL Patent: WO 0100805-A 621 04-JAN-2001;
          BASF AKTIENGESSELLSCHAFT (DE)
FEATURES
  source
    1. .1335
      /location/Qualifiers
      /organism="Corynebacterium glutamicum"
      /mol_type="unassigned DNA"
      /db_xref="taxon:1718"
      101. .1315
      /notes="unnamed protein product; FRXA02233"
      /codon_start=1
      /transl_table=11
      /protein_id="CAC26747.1"
      /db_xref="GI:12544748"
      /db_xref="RENTREMBL:CAC26747"
      /translation="MLVSTGWTVHGDGKIEPGAVAPKRLSGWRTIGIGMORHVI
      AMFGATLLVPTLTGFPVNTTLFSGLGTLILITNRLPSYLGSSFAFIAPLTATQV
      HGIQVOIGILVAGLVVAIGFVWKAQKRVDAVMPAVTGAIVALLIGLAPTAAG
      NFSSQPLVATATLPAIIATVAGSGMIAPLGLIGVIGVFAITNLSEGAADTIR
      EAAWPGLPFHKEPQLSAILLVLPVILVIAENVGHVXAVSEMTEGDDLDLAGDALI
      ADGGFTLAGAGGSGTTIAENIGVNAATRVSTAAVYVAACTAIALAFIKFGALI
      FTIPAGVLGACILVLYGLIMGLIRIQDNKVNFPNLTMAAVALVAGIGNLTIV
      FGVTLLEGIAWAL"
ORIGIN
  Query Match 66.7%; Score 18; DB 6; Length 1335;
  Best Local Similarity 80.8%; Pred. No. 8.4e+02;
  Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 CACCCGCGCTTTAGGATCGACACCTG 26
    |||||
Db 190 CAGGCGCTCTTTAGGAGCGACACTG 165

RESULT 18
AX067037/c
LOCUS AX067037 1410 bp. DNA linear PAT 24-JAN-2001
DEFINITION Sequence 619 from Patent WO0100805.
ACCESSION AX067037
VERSION AX067037.1 GI:12544745
KEYWORDS Corynebacterium glutamicum
SOURCE Corynebacterium glutamicum
ORGANISM Corynebacterium glutamicum
REFERENCE 1
AUTHORS Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.
TITLE Corynebacterium glutamicum genes encoding proteins involved in
membrane synthesis and membrane transport
JOURNAL Patent: WO 0100805-A 619 04-JAN-2001;
          BASF AKTIENGESSELLSCHAFT (DE)
FEATURES
  source
    1. .1410
      /location/Qualifiers
      /organism="Corynebacterium glutamicum"
      /mol_type="unassigned DNA"
      /db_xref="taxon:1718"
      101. .1390
      /note="unnamed protein product; RXN02233"
      /codon_start=1
      101. .1390
      /note="unnamed protein product; RXN02233"
      /codon_start=1
  
```



LOCUS AE009079 11198 bp DNA linear BCT 20-DEC-2001  
DEFINITION Agrobacterium tumefaciens strain C58 circular chromosome, section  
105 of 256 of the complete sequence.  
ACCESSION AE009079 AE008688  
VERSION AE009079.1 GI:17739557

Zhou, Y., Bovee Sr., D., Clendenning, J., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutayavin, T., Levy, R., Li, M., McCalland, E., Palmieri, A., Raymond, C., Rouse, G., Saenichimachak, C., Wu, Z., Gordon, D., Eisen, J. A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S. V., Tomb, J., Gordon, M. P., Olsson, M. V., and Westergaard, F. W.

MEDLINE  
 PUBMED  
 21608550  
 11743193  
 2 (bases 1 to 11198)  
 Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,  
 Chen, Y., Wool, L., Kitajima, T., Okura, Y., Almeida, T., N.E.

Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutayavin, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J. A., Paulsen, I., Ram, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, D., Yang, M.

Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S. V., Tomb, J., Gordon, M. P., Olson, M. V. and Nester, F. W.

**TITLE**  
JOURNAL  
Direct Submission  
Submitted (27-SEP-2001) Department of Microbiology, University of  
Washington, 1959 NE Pacific Ave. Box 357242, Seattle, WA  
Data Recd., Date:

FEATURES	source	Location/Qualifiers
1. .ll198		
		/organism="Agrobacterium tumefaciens str. C58 (U. Washington)"
		(see Agrobacterium tumefaciens)

```

/strain="C58"
/db_xref="taxon:180835"
135..1616
gene
/gene="Atu1156"
135..1616
CDS
135..1616

```

```

/ncore=#identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=1
/product="sulfate permease"
/protein_id="AAI42169.1"
/db_xref="GI:1733558"

```

IAAGFAKGLGVRMEFVSVMGTFNAILIEMAQGLPELIGVPVETVYNIAAGLAIIT  
LPLVTKIIPSPULVAIVALTQCVAFVSGMDIRTVGDLGELPSTLPIFALPQVPLFTFL  
QIIFPYSIALAAVGLBESLLTAQIVDDMTDTSNKSQECIGQGTNSIASALIGMGCC

gene  
KVVLFKFKTKGRVIGVNEASAHMIDRFALHDKPDAAAAPLH"  
complement (1657 2847)  
PELSADGRIRYVDGQIFFASTESFAAFDFAEEVEAVTIDVTQHLWDISAVGALD

```
/gene= ACU113/  
/note="identified by sequence similarity; putative; ORF
```

located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AA42170.1"  
 /db\_xref="GI:117739559"  
 /translation="MRLKSTRQVLFAAALLTLTSAFALPDGDVLLKTKNNAAYAAQQG  
 EIAAGSVANGDVTWVTFNFAADAPAKKI PVGNITLEGVEENNGYKIKNARFDNI  
 DYQKEVLIKASDINYNSGLVIPADATVYNDALMFVDEAHSGPI SVSYDQKAFLEE  
 TSTAKMSVDKNSIGSELVDSGFKADLSEVTDPAIKDAIEKLKQLAOLSDMTMSQSGE  
 VASGTVDERIAYLIDDDVGLNMSFGPSGYTLAFIKSMOSTVKAQANPKCAQDAAS  
 SLAMGLUMQNSGFVNABEISFEDASIKTRALDIYAASQGMGTGQLAQTIKGMAPIMKAS  
 LNVPELQNNVSTAINAYIDNPKNFSITATSEKPEKVPFFFMINGAAMGAPNTLTKMLGVTV  
 TANE"  
 3077. 5329  
 /gene="parC"  
 /note="synonym: Atu1158"  
 3077. 5329  
 /gene="parC"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=11  
 /product="topoisomerase IV subunit A"  
 /protein\_id="AA42171.1"  
 /db\_xref="GI:17739560"  
 /translation="MGKDLVPPSGDDNIHPVLKKAALBERYLAYALSTIMHRAFLDQ  
 RDGLKPVHRRIIHAFSEMGIRPNSAFKKCARI GVDVIGKPHPGDQSDVALVRLPAQ  
 FSRQYPVDDQNGFNITDGCAGAAFTYKARTMDVAALLEGIGEDAVDFPRTYNEED  
 EEPVLPCAPNLILANGASGTAVGMTSIPHNABELCDAAHLIKHPDATVEKLFVE  
 IPGDLPITGGVITISERNLIDAYVTGRGGRVRAKWTEDLGRGGYQIVITETIPQVQ  
 KSRLIEAELLIAKPLLEDVDESAEDVRVLPVKNRTVDATLLMESLPLRLSLE  
 SRYLPNNVLSLGVKPKVMAINVELSEWLAHRKDVLAHRSRHRLLAIDRELETLGLGL  
 IAYLNDEVRIRIEEEDPQVMAWLSLTDNCAEAILNMLRLNLEKLEFEPIRKEHD  
 DLSKEIEGLLSDQKQWTVAMEIGEVKKYAKATEIGRRRTQFADAPDAIDAEI  
 QQAMIKEKPVITVTSQKGMIRALKGHMSDTSALTFFEGDGPKLAFPAQITDKULLLIT  
 GKGYATLGADLKPGCRGHGPIRIMVMDENDQDILAFVHPDPSRKILLVSTGQVEMGR  
 AESMIANTKRGKIMVNSMPDEAKLAVPTVGDHVAVGENRKLLAFPLAQVQPMERG  
 KGVRLQRYKGGVVDVKCFALADGLSNSTAGRAFTKVGEELEWLADRAVTGRTVPK  
 GFKPSGKFGG"  
 complement(5426. 6184)  
 /gene="Atu1159"  
 complement(5426. 6184)  
 /gene="Atu1159"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=11  
 /product="arginine-tRNA-protein transferase"  
 /protein\_id="AAL42172.1"  
 /db\_xref="GI:17739561"  
 /translation="MNTQATSPQFYLTAPATCYLPNQMKVFTHLVGPRAPEMND  
 LLTGQFRRSNIAYRACETCRACVSRILTEQPTKRCRVRVLAANSVAVATVHA  
 EPSQEAFALFRYLDHFRHQSMDMSALDYAIWVEDVTNTRIIIEVRPEQSGSFI  
 SKRGELLVALVSDVMSDGLSNVSYFFNPELEKSLGTFTMLIDHITRTRALGLPHVYL  
 YWVGSEKMGYKTRYHPOEHLTPRGWEIYSPKE"  
 complement(6362. 6820)  
 /gene="Atu1160"  
 complement(6362. 6820)  
 /gene="Atu1160"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAL42173.1"  
 /db\_xref="GI:17739562"  
 /translation="MSLSQNTPYAAPDDWRAYSGVLSRRIFAFILIDYAILLLICIPAA  
 VIVFLGVITLGLHFLPALFVIVAVLYFGVSLGSPQATPGMRAMGIAMRMDGR  
 IDFLALVTHI VLFVWVINSVLPILLAGLFTERSRLVHDFLLGLTVITR"  
 complement(6918. 7931)  
 /gene="hemB"  
 /note="synonym: Atu1161"  
 3077. 5329  
 /gene="parC"  
 /note="synonym: Atu1158"  
 3077. 5329  
 /gene="parC"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAL42174.1"  
 /db\_xref="GI:17739563"  
 /translation="MSLSQNTPYAAPDDWRAYSGVLSRRIFAFILIDYAILLLICIPAA  
 VIVFLGVITLGLHFLPALFVIVAVLYFGVSLGSPQATPGMRAMGIAMRMDGR  
 IDFLALVTHI VLFVWVINSVLPILLAGLFTERSRLVHDFLLGLTVITR"  
 complement(6918. 7931)  
 /gene="hemB"  
 /note="synonym: Atu1161"

```

CDS
complement (6918. 79311)
/gene="hems"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="delta-aminolevulinic acid dehydratase"
/protein_id="AAL42174.1"
/db_xref="GI:17739563"
/translation="MQKTHLVSDVITGRRMRNRKADWTRRLVQENRLTVDDLKPV
FVPSQNVETIPAMPGVNMSVDRVFAAKEADIGIPAIATFPNIMELRDETSN
ALEANNLNQIATKAVKAVNIGITDVALDPTFTSHGDIUGDBDIVNDETVEQVT
AAILAGSAGSDIISPEMDRGICATIRGLDAAHQNVIMAYATKFSGGYGYPYREA
ISTAGLLGDKNSYIVSPANGMEAIRDAALDVEEGADMVMKPLPYLIDICWRLKENF
GIPTAYQVSGEYTIQIKAAANGWIDGVRVWMTLLCFKRAAGCDGILTYFAIEAAARKL
AKH"
gene
8057. 8542
/gene="Atu1162"
CDS
8057. 8542
/gene="Atu1162"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL42175.1"
/db_xref="GI:17739564"
/translation="MOYGLVAPQSRPYLPAMEPQSPKRPRLTEILFLVFMRLVAVS
CFWPGLOYAMITGVSLGCHGRFQMLNPMRWAGSALAVIPVAALGLWLGASWGAVM
WYMGAGQTAMKXVPHFEGATLVFVWHGLVATYILFLVAFWLDHRQNEERARDL
P"
gene
8774. 9286
/gene="mucs"
/notes="synonym: Atu1163"
CDS
8774. 9286
/gene="mucs"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="exopolysaccharide II synthesis transcriptional
activator ExpG"
/protein_id="AAL42176.1"
/db_xref="GI:17739565"
/translation="MINSKVPQAVDAHEDTTRLSYMSLHVLVERHLRDLVLIK
EFDQRDDVNAVQALLFNIGNSLETAGEURSGYGLGNSVKNVKLVLDGLGQWAGH
RSFQDVSRSYRIGLITBEGQAIETVARLYERHVSIEKVGIGTGEFSEMMKLLQRLDR
FWDNSIAYRL"
gene
9660. 10991
/gene="Atu1164"
CDS
9660. 10991
/gene="Atu1164"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"

Query Match 66.7%; Score 18; DB 1; Length 11198;
Best Local Similarity 80.8%; P: 0.0; Mismatches 5; Indels 0; Gaps 0;
Conservative 30.0%;

```

Qy  
Db

1 CACCCGGCGCTTTAGGATCGACACCTG 26  
||| ||| ||| ||| ||| ||| ||| |||  
278 CAGCCCCCACTTTCGGATCGACACCGG 253

RESULT 21  
AE008044/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE

AE008044 14354 bp DNA linear BCT 18-DEC-2000  
Agrobacterium tumefaciens str. C58 circular chromosome, section 102  
of 254 of the complete sequence.  
AE008044 AE007869  
AE008044.1 GI:15156183  
Agrobacterium tumefaciens str. C58 (Cereon)





/note="Sequence from overlapping clone RP11-311B18  
(AL139818). Assembly confirmed by restriction digest."

## misc\_feature

56166..56252

/note="match: STS: Em:A0048351"

## misc\_feature

57887..58177

/note="match: STS: Em:G50546"

## ORIGIN

Query Match 66.7%; Score 18; DB 9; Length 84615;  
Best Local Similarity 80.8%; Pred. No. 7.8e+02;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTG 26

Db 29570 CACCTGTGCTTCAGGATCCTCACCTG 29545

## RESULT 23

AL954650

LOCUS

DEFINITION Human DNA sequence from clone RP11-523M19 on chromosome 1, complete

ACCESSION

AL954650

VERSION

AL954650.8

KEYWORDS

HTG.

SOURCE

ORGANISM

Homo sapiens (human)

REFERENCE

1 (bases 1 to 86257)

AUTHORS

Martin, S.

TITLE

JOURNAL

COMMENT

Submitted (25-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgehire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk

On Feb 25, 2003 this sequence version replaced gi:28208053.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

-----

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the rare

assembly was confirmed by restriction digest, except on the rare

occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information

on the WORMPEP database can be found at

[http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence

was generated from part of bacterial clone contigs of human

chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping

Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr1>

RP11-523M19 is from the library RP11-11.2 constructed by the group

of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

Location/Qualifiers

1..86257

/organism="Homo sapiens"

/mol\_type="genomic DNA"

FEATURES

source

## ORIGIN

Query Match 66.7%; Score 18; DB 9; Length 86257;

Best Local Similarity 80.8%; Pred. No. 7.8e+02;

Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTG 26

Db 5930 CACCCGCGCTTCAGGATCCTCACCTG 5955

## RESULT 24

AL954830

LOCUS

DEFINITION

Mouse DNA sequence from clone RP23-12A10 on chromosome 11, complete

sequence.

AL954830

VERSION

AL954830.7

KEYWORDS

HTG.

SOURCE

ORGANISM

Mus musculus (house mouse)

REFERENCE

1 (bases 1 to 109889)

AUTHORS

Matthews, L.

TITLE

JOURNAL

COMMENT

Submitted (10-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgehire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk

On Dec 16, 2002 this sequence version replaced gi:26788337.

Sequence from the Mouse Genome Sequencing Consortium whole genome

shotgun may have been used to confirm this sequence. Sequence data

from the whole genome shotgun alone has only been used where it has

a phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

-----

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the rare

assembly was confirmed by restriction digest, except on the rare

occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information

on the WORMPEP database can be found at

[http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence

was generated from part of bacterial clone contigs of human

chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping

Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr1>

RP23-12A10 is from the library RP23-12A10 constructed by the group

of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

Location/Qualifiers

1..109889

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="11"

## FEATURES

source

```
ORIGIN
/clone="RP23-12A10"
/clone_lib="RPCI-23"

Query Match      66.7%; Score 18; DB 10; Length 109889;
Best Local Similarity 80.8%; Pred. No. 7.8e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTG 26
Db 25110 CTCCTAGCTTCAGATCCACACCTG 25135

RESULT 25
LMFLCHR16_02/c
WPCOMMENT
Sequence split into 11 fragments LOCUS LMFLCHR16 Accession AL499619
Fragment Name      Begin      End
LMFLCHR16_00      1      110000
LMFLCHR16_01      100001  210000
LMFLCHR16_02      200001  310000
LMFLCHR16_03      300001  410000
LMFLCHR16_04      400001  510000
LMFLCHR16_05      500001  610000
LMFLCHR16_06      600001  710000
LMFLCHR16_07      700001  810000
LMFLCHR16_08      800001  910000
LMFLCHR16_09      900001  1010000
LMFLCHR16_10     1000001  1030105
Continuation (3 of 11) of LMFLCHR16 from base 200001 (AL499619 Leishmania major chromosome)

Query Match      66.7%; Score 18; DB 2; Length 110000;
Best Local Similarity 80.8%; Pred. No. 7.8e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTG 26
Db 106392 CGCCGCGCTCGGATCGACACCCG 106367

RESULT 26
LMFLCHR16_03/c
WPCOMMENT
Sequence split into 11 fragments LOCUS LMFLCHR16 Accession AL499619
Fragment Name      Begin      End
LMFLCHR16_00      1      110000
LMFLCHR16_01      100001  210000
LMFLCHR16_02      200001  310000
LMFLCHR16_03      300001  410000
LMFLCHR16_04      400001  510000
LMFLCHR16_05      500001  610000
LMFLCHR16_06      600001  710000
LMFLCHR16_07      700001  810000
LMFLCHR16_08      800001  910000
LMFLCHR16_09      900001  1010000
LMFLCHR16_10     1000001  1030105
Continuation (4 of 11) of LMFLCHR16 from base 300001 (AL499619 Leishmania major chromosome)

Query Match      66.7%; Score 18; DB 2; Length 110000;
Best Local Similarity 80.8%; Pred. No. 7.8e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTG 26
Db 6392 CGCCGCGCTCGGATCGACACCCG 6367

RESULT 27
AC122465
LOCUS
DEFINITION Mus musculus BAC clone RP24-304r22 from chromosome 13, complete sequence.
ACCESSION AC122465
```

AC122465.4 GI:26024143  
HTG.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 142340)  
Holmes, A., Haglund, K. and Dignan, G.  
The sequence of Mus musculus BAC clone RP24-304r22  
Unpublished (2001)  
2 (bases 1 to 142340)  
Wilson, R.  
Sequencing of Mus musculus  
Unpublished (2001)  
3 (bases 1 to 142340)  
McPherson, J.D. and Waterston, R.H.  
Direct Submission  
Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
4 (bases 1 to 142340)  
McPherson, J.D. and Waterston, R.H.  
Direct Submission  
Submitted (13-NOV-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
5 (bases 1 to 142340)  
McPherson, J.D. and Waterston, R.H.  
Direct Submission  
Submitted (04-DEC-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
6 (bases 1 to 142340)  
Wilson, R.  
Direct Submission  
Submitted (13-NOV-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Dec 4, 2002 this sequence version replaced gi:24943033.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Summary Statistics  
-----  
Center project name: M\_BB0304122

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:  
The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:  
This sequence is the entire insert of the clone.  
Location/Qualifiers  
1. 142340  
source

```

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="13"
/map="13"
/clone="RP24-304122"
/clone_lib="RPC1-24"
350..419
/rpt_family="ID"
429..535
/rpt_family="Alu"
941..975
/rpt_family="tRNA-Phe-TTY"
947..1128
/rpt_family="B4"
2257..2449
/rpt_family="B2"
2468..2578
/rpt_family="L1"
2648..2788
/rpt_family="Alu"
2889..3019
/rpt_family="Alu"
4711..4793
/rpt_family="ID"
4794..4831
/rpt_family="B4"
4869..4984
/rpt_family="Alu"
5046..5346
/rpt_family="L1"
5746..5836
/rpt_family="Alu"
6301..6337
/rpt_family="CR1"
6651..6782
/rpt_family="Alu"
6874..7003
/rpt_family="MaLR"
7472..7664
/rpt_family="B2"
7713..7826
/rpt_family="MIR"
7903..8227
/rpt_family="MaLR"
8404..8648
/rpt_family="B4"
9311..9449
/rpt_family="Alu"
9573..9673
/note="Sequence derived from one plasmid subclone."
10586..10755
/rpt_family="L1"
10843..10937
/rpt_family="ID"
11030..11146
/rpt_family="L1"
11147..11508
/rpt_family="MaLR"
11509..11602
/rpt_family="L1"
12000..12030
/rpt_family="B2"
12031..12189
/rpt_family="B4"
12186..12396
/rpt_family="MaLR"
12714..12878
/rpt_family="ERV1"
12879..13001
/rpt_family="Alu"
13139..13390
/rpt_family="ERV1"

```

```

repeat_region 13391..13574
/rpt_family="B2"
repeat_region 13575..13689
/rpt_family="ERV1"
repeat_region 15299..15355
/rpt_family="ID"
repeat_region 15356..15459
/rpt_family="B4"
repeat_region 16335..16412
/rpt_family="B4"
repeat_region 16413..16556
/rpt_family="Alu"
repeat_region 16560..16705
/rpt_family="Alu"
repeat_region 16706..16732
/rpt_family="BC1_MX"
repeat_region 17753..17785
/rpt_family="B2"
repeat_region 17892..17939
/rpt_family="ERV1"
repeat_region 18684..18772
/rpt_family="Alu"
repeat_region 18826..19103
/rpt_family="B4"
repeat_region 18894..19950
/rpt_family="L1"
repeat_region 19964..20066
/rpt_family="Alu"
repeat_region 20631..20971
/rpt_family="MaLR"
repeat_region 20978..21471
/rpt_family="L1"
repeat_region 23096..23124
/rpt_family="B4"
repeat_region 23125..23250
/rpt_family="Alu"
repeat_region 23262..23724
/rpt_family="L1"
repeat_region 23732..24165
/rpt_family="RMR10A"
repeat_region 24219..24298
/rpt_family="ID"
repeat_region 24316..24385
/rpt_family="B4"
repeat_region 24658..24768
/rpt_family="Alu"
repeat_region 24895..24947
/rpt_family="ERV1"
repeat_region 25277..25504
/rpt_family="B4"

```

Query Match 66.7%; Score 18; DB 10; Length 142340;

Best Local Similarity 80.8%; Pred. No. 7.7e+02; Mismatches 5; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTAGGATCGACACCTG 26

Db 63795 CACCCGCGCTTAGGCTCCACATG 63820

RESULT 28

AC106319/c

LOCUS

DEFINITION

AC106319

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 184424)

AC106319 184424 bp DNA linear HTG 13-MAY-2003

Rattus norvegicus clone CH230-239A7, WORKING DRAFT SEQUENCE.

AC106319 HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

AC106319.4 GI:30581450

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

AUTHORS	Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulsegged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Staimle, M., Strong, R., Sutton, A., Svatek, A., Tabok, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
	Direct Submission																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
	Unpublished																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
	2 (bases 1 to 184424)																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
	Worley, K.C.																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
	Direct Submission																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
	Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
	3 (bases 1 to 184424)																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
	Rat Genome Sequencing Consortium.																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
	Direct Submission																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
On May 13, 2003 this sequence version replaced gi:23608511.																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas ( <a href="http://www.hgsc.bcm.tmc.edu/projects/rat/">http://www.hgsc.bcm.tmc.edu/projects/rat/</a> ). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
REFERENCE																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														



Miller, B., Nascimento, L. U., O'Shaughnessy, A. L., Preston, R. R., Rodriguez, S., Santos, L., Shah, R. S., Spiegel, L. A., Toch, K., Vill, M. D. and Zucaverni, F.

# Mouse Genomic Sequence

## JOURNAL

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

Unpublished

2 (bases 1 to 206112)

McCombie, W. R.

Direct Submission

Submitted (01-NOV-2000) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

### COMMENT

----- Genome Center  
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratory

Center code: CSHL

Web site: <http://www.cshl.org/genseq>

Contact: [mccombie@cshl.org](mailto:mccombie@cshl.org)

----- Project Information

Center project name: RP23-174G20

Center clone name: RP23-174G20

\*\*\* NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 17043: contig of 17043 bp in length
1 17214: gap of unknown length
17215 32036: contig of 14822 bp in length
32037 32077: gap of unknown length
32208 46766: contig of 14559 bp in length
46767 46937: gap of unknown length
46938 57574: contig of 10637 bp in length
57575 57745: gap of unknown length
57746 68166: contig of 10421 bp in length
68167 68337: gap of unknown length
68338 77389: contig of 9052 bp in length
77390 77560: gap of unknown length
77561 86460: contig of 8900 bp in length
86461 86630: gap of unknown length
86631 94462: contig of 7832 bp in length
94463 102252: gap of unknown length
102253 102422: gap of unknown length
102423 109611: contig of 7189 bp in length
109612 109781: gap of unknown length
109782 116602: contig of 6821 bp in length
116603 116773: gap of unknown length
116773 123357: contig of 5585 bp in length
123358 123527: gap of unknown length
123528 123808: contig of 6281 bp in length
123809 129978: gap of unknown length
129979 135993: contig of 6015 bp in length
135994 136163: gap of unknown length
136164 141952: contig of 5788 bp in length
141952 142122: gap of unknown length
142122 147858: contig of 5737 bp in length
147859 148028: gap of unknown length
148029 153410: contig of 5382 bp in length
153411 153580: gap of unknown length
153581 153783: contig of 5203 bp in length
153784 158953: gap of unknown length
158954 163793: contig of 4840 bp in length
163794 163794: gap of unknown length
163794 163964: contig of 4834 bp in length
163965 168977: gap of unknown length
168978 173706: contig of 4739 bp in length
173707 173877: gap of unknown length
173878 177928: contig of 4052 bp in length
177929 178098: gap of unknown length

```

```

* 178099 181857: contig of 3759 bp in length
* 181858 182027: gap of unknown length
* 182028 184932: contig of 2905 bp in length
* 184933 185102: gap of unknown length
* 185103 187993: contig of 2891 bp in length
* 187994 188163: gap of unknown length
* 188164 191028: contig of 2865 bp in length
* 191029 191198: gap of unknown length
* 191199 193996: contig of 2798 bp in length
* 193997 194166: gap of unknown length
* 194167 196953: contig of 2787 bp in length
* 196954 197123: gap of unknown length
* 197124 199647: contig of 2524 bp in length
* 199648 199817: gap of unknown length
* 199818 202135: contig of 2318 bp in length
* 202136 202305: gap of unknown length
* 202306 204589: contig of 2284 bp in length
* 204590 204759: gap of unknown length
* 204760 206112: contig of 1353 bp in length.

```

### FEATURES

Location/Qualifiers  
1..206112  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP23-174G20"

### ORIGIN

```

Query Match      66.7%; Score 18; DB 2; Length 206112;
Best Local Similarity 80.8%; Pred. No. 7.7e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTGA 27
      ||||| ||||| ||||| ||||| |||||
Db 74245 ACCCGCGCTTCTGATCTACACGAGA 74220

AC099282      213462 bp DNA linear HTG 21-SEP-2002
Rattus norvegicus clone CH230-130019, *** SEQUENCING IN PROGRESS
*** 4 unordered pieces.
AC099282      GI:23265682
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 213462)
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Y., Chen, Y., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draber, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, N.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpthy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

```

```

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensheva, L., Loulsegged, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, K., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwunonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasceriak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 213462)
Worley, K.C.
Direct Submission
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 213462)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 21, 2002 this sequence version replaced gi:21795359.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GKCN
Center clone name: CH230-130019
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 208516 bases at least Q40
Consensus quality: 209366 bases at least Q30
Consensus quality: 210059 bases at least Q20
Estimated insert size: 227892; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

```

* be preserved.
* 1 166582: contig of 166582 bp in length
* 166583 166682: gap of unknown length
* 166683 168153: contig of 1471 bp in length
* 168154 168253: gap of unknown length
* 168254 183417: contig of 15164 bp in length
* 183418 183517: gap of unknown length
* 183518 213462: contig of 29945 bp in length.
* Location/Qualifiers
* 1..213462
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-130019"
* complement(163258..164086)
* /note="clone boundary"
* clone_end:T7
* site:EcoRI
* end_sequence:BH342154"

ORIGIN
Query Match 66.7%; Score 18; DB 2; Length 213462;
Best Local Similarity 80.8%; Pred. No. 7.7e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTG 26
Db 56485 CACCCGCGCTTTAGGATCGACACCTG 56460

Search completed: June 20, 2004, 11:42:32
Job time : 760.143 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:46:26 ; Search time 155.143 seconds  
(without alignments)  
739.327 Million cell updates/sec

Title: US-10-624-714-12  
Perfect score: 27  
Sequence: 1 caccgcgcttaggcacacotga 27

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 300 summaries

Database : N Geneseq 29Jan04:\*

1: Geneseqn1980s: \*  
2: Geneseqn1990s: \*  
3: Geneseqn2000s: \*  
4: Geneseqn2001as: \*  
5: Geneseqn2001bs: \*  
6: Geneseqn2002as: \*  
7: Geneseqn2003as: \*  
8: Geneseqn2003bs: \*  
9: Geneseqn2003cs: \*  
10: Geneseqn2004as: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	27	100.0	110000	4	AAI99682 25
C 2	27	100.0	110000	4	AAI99682 26
C 3	27	100.0	110000	4	AAI99683 25
C 4	27	100.0	110000	4	AAI99683 26
C 5	18.6	68.9	2000	6	ABZ16792
C 6	18	66.7	1287	5	AAH67093
C 7	18	66.7	1335	4	AAH68053
C 8	18	66.7	1410	4	AAH68052
C 9	18	66.7	34980	5	AAH68330
C 10	17.6	65.2	766	4	AAH68335
C 11	17.6	65.2	771	4	AAH68335
C 12	17.6	65.2	894	7	AAH68335
C 13	17.6	65.2	933	7	AAH68335
C 14	17.6	65.2	1195	6	AAH68335
C 15	17.6	65.2	4919	6	AAH68335
C 16	17.4	64.4	1789	7	AAH68347
C 17	17.4	64.4	2283	4	AAH68347
C 18	17.4	64.4	2283	7	AAH68347
C 19	17.4	64.4	2937	4	AAH68347
C 20	17.4	64.4	5484	4	AAH68347
C 21	17.4	64.4	110000	4	AAI99682 40
C 22	17.4	64.4	110000	4	AAI99683 40
C 23	17.2	63.7	999	6	ABK74646

24	17.2	63.7	1257	7	ACA36049
C 25	17	63.0	279	4	AAK44543
C 26	17	63.0	279	4	ABS44201
C 27	17	63.0	279	6	ABS18780
C 28	17	63.0	349	6	ABS69598
C 29	17	63.0	463	4	AAK31460
C 30	17	63.0	463	4	ABS31142
C 31	17	63.0	463	6	ABS06214
C 32	17	63.0	618	3	AAA44965
C 33	17	63.0	897	7	ACA28624
C 34	17	63.0	1485	7	ACA20990
C 35	17	63.0	1674	7	ACA36278
C 36	17	63.0	1675	7	ACA32299
C 37	17	63.0	1701	2	AAQ65617
C 38	17	63.0	1701	2	AAV64813
C 39	17	63.0	1701	2	AAZ22736
C 40	17	63.0	1701	3	AAZ33270
C 41	17	63.0	1701	3	AAZ95673
C 42	17	63.0	1701	3	AAZ46281
C 43	17	63.0	1701	3	AAZ37820
C 44	17	63.0	12409	4	AAZ30238
C 45	17	63.0	12409	7	ACD01493
C 46	17	63.0	12838	2	AAZ20509
C 47	17	63.0	17736	4	ABL28554
C 48	16.8	62.2	8163	4	AAK84484
C 49	16.8	62.2	8163	4	AAK84484
C 50	16.6	61.5	467	6	ABL93494
C 51	16.6	61.5	546	6	ABL2882
C 52	16.6	61.5	587	2	AAI19928
C 53	16.6	61.5	920	7	ABZ51070
C 54	16.6	61.5	939	7	ACA40224
C 55	16.6	61.5	1357	5	AAZ85511
C 56	16.6	61.5	1467	3	AAZ32779
C 57	16.6	61.5	1587	7	ABX78350
C 58	16.6	61.5	1675	7	ACA49450
C 59	16.6	61.5	1677	7	ACA51446
C 60	16.6	61.5	1712	5	AAZ77240
C 61	16.6	61.5	1901	3	AAZ9864
C 62	16.6	61.5	1995	7	ACA23852
C 63	16.6	61.5	2218	5	AAZ74300
C 64	16.6	61.5	2575	5	AAZ71248
C 65	16.6	61.5	3164	4	ABL25304
C 66	16.6	61.5	3211	5	AAZ77492
C 67	16.6	61.5	5927	4	ABL6006
C 68	16.6	61.5	5927	4	AAZ57148
C 69	16.6	61.5	5927	9	ADC35863
C 70	16.4	60.7	100	7	ACD80473
C 71	16.4	60.7	102	2	AAQ70234
C 72	16.4	60.7	258	3	AAI15651
C 73	16.4	60.7	318	2	AAQ12499
C 74	16.4	60.7	423	4	AAI93517
C 75	16.4	60.7	427	2	AAQ70220
C 76	16.4	60.7	438	3	AAI15656
C 77	16.4	60.7	555	2	AAQ03732
C 78	16.4	60.7	555	2	AAQ02288
C 79	16.4	60.7	595	4	AAI19308
C 80	16.4	60.7	595	4	ABA31457
C 81	16.4	60.7	595	4	AAK38507
C 82	16.4	60.7	595	4	AAK12784
C 83	16.4	60.7	633	1	AAZ92729
C 84	16.4	60.7	698	2	AAQ10159
C 85	16.4	60.7	723	9	ADC92366
C 86	16.4	60.7	735	6	ABK74283
C 87	16.4	60.7	823	2	AAQ13323
C 88	16.4	60.7	902	2	AAQ81202
C 89	16.4	60.7	951	2	AAQ11967
C 90	16.4	60.7	975	7	ACA51639
C 91	16.4	60.7	1004	7	ACA49007
C 92	16.4	60.7	1011	7	ACA31854
C 93	16.4	60.7	1153	2	AAQ92410
C 94	16.4	60.7	1188	2	ACA44613
C 95	16.4	60.7	1368	2	AAV04530
C 96	16.4	60.7	1742	5	AAZ89852

ACA36049	Prokaryot
AAK44543	Human bon
ABS44201	Human liv
ABS18780	Human gen
ABS69598	Novel mur
AAK31460	Human bon
ABS31142	Human liv
ABS06214	Human gen
AAA44965	Human sec
ACA28624	Prokaryot
ACA20990	Prokaryot
ACA36278	Prokaryot
ACA32299	Prokaryot
AAQ65617	Human zon
AAV64813	Human ZPB
AAZ22736	Human zon
AAZ33270	Human zon
AAZ95673	Human zon
AAZ46281	Human ooc
AAZ37820	Human zon
AAZ30238	DNA encod
ACD01493	Human gen
AAZ20509	Polynucle
ABL28554	Drosophil
AAK84484	Human imm
AAI05749	Human rep
ABL93494	Arabidops
ABL2882	Arabidops
AAI19928	Lily male
ABZ51070	Aspergill
ACA40224	Prokaryot
AAZ85511	DNA encod
AAZ32779	Arabidops
ABX78350	Wheat str
ACA49450	Prokaryot
ACA51446	Prokaryot
AAZ77240	DNA encod
AAZ9864	Human sec
ACA23852	Prokaryot
AAZ74300	DNA encod
AAZ71248	DNA encod
ABL25304	Drosophil
AAZ77492	DNA encod
ABL16006	Drosophil
AAZ57148	DNA encod
ADC35863	Drosophil
ACD80473	E. coli K
AAQ70234	HTLV-1 3A3
AAI15651	HTLV gp21
AAQ12499	Nucleotid
AAI93517	Human pol
AAQ70220	HTLV p21
AAI15656	HTLV anti
AAQ03732	Gene enco
AAQ02288	Human T-c
AAI19308	Probe #92
ABA31457	Probe #99
AAK38507	Human bon
AAK12784	Human bra
AAZ92729	HIV-1 - h
AAQ10159	Sequence
ADC92366	E. faeciu
ABK74283	Bacillus
AAQ13323	HTLV-1 en
AAQ81202	Cloned al
AAQ11967	HIV-1/HTL
ACA51639	Prokaryot
ACA49007	Prokaryot
ACA31854	Prokaryot
AAQ92410	Escherich
ACA44613	Prokaryot
AAV04530	HTLV-I 5'
AAZ89852	DNA encod

97	16.4	60.7	1797	2	AQ011368	Aac11968	HIV-2/HIV
98	16.4	60.7	2095	4	ABL13403	Abi13403	Drosophila
99	16.4	60.7	2165	2	AAT59702	Aat59702	Endotoxin
100	16.4	60.7	2220	1	AAN91238	Aan91238	Bi antige
101	16.4	60.7	2331	6	ACA24911	Act24911	Prokaryot
102	16.4	60.7	2565	6	ABT10715	Abt10715	Human bre
103	16.4	60.7	2776	9	AD224910	Ad224910	DNA encod
104	16.4	60.7	2865	6	ABN97212	Abn97212	Gene #371
105	16.4	60.7	2761	9	AD285176	Ad285176	Human pro
106	16.4	60.7	2761	9	AD285176	Ad285176	Farnesyl
107	16.4	60.7	2865	4	AAK52354	Aak52354	Human pol
108	16.4	60.7	2919	4	ABL12405	Abi12405	Drosophila
109	16.4	60.7	2951	4	AAK53338	Aak53338	Human pol
110	16.4	60.7	3054	5	AAK585672	Aak585672	DNA encod
111	16.4	60.7	3324	4	ABU17295	Abu17295	Drosophila
112	16.4	60.7	4303	5	AAK585519	Aak585519	DNA encod
113	16.4	60.7	4692	4	ABL13402	Abi13402	Drosophila
114	16.4	60.7	5262	4	ABL12404	Abi12404	Drosophila
115	16.4	60.7	5813	4	AAH57490	Aah57490	Human liv
116	16.4	60.7	7095	4	ABU17294	Abu17294	Drosophila
117	16.4	60.7	7287	4	AAH57428	Aah57428	Human int
118	16.4	60.7	9045	2	AAAT42902	Aat42902	DNA sequ
119	16.4	60.7	9047	1	AAAN40080	Aan40080	Sequence
120	16.4	60.7	13852	5	ABA17744	Abal7744	Human ner
121	16.4	60.7	13852	5	ABA17745	Abal7745	Human ner
122	16.4	60.7	13852	5	ABA21396	Abal21396	Human ner
123	16.4	60.7	13852	5	ABA21395	Abal21395	Human ner
124	16.4	60.7	18776	4	ABU16312	Abu16312	Drosophila
125	16.4	60.7	23078	4	AAK595508	Aak595508	Propionib
126	16.4	60.7	23078	4	ACF64437	Act64437	Propionib
127	16.4	60.7	65011	6	ADA02831	Ada02831	Mouse Mef
128	16.4	60.7	96599	9	ADK72569	Adk72569	Mouse Mef
129	16.4	60.7	96599	9	ADK72569	Adk72569	Mouse Mef
130	16.4	60.7	96599	9	ADK72569	Adk72569	Mouse Mef
131	16.4	60.7	110000	2	AAAT42902	Aat42902	DNA encod
132	16.4	60.7	110000	4	AAI99682	Aai99682	DNA encod
133	16.4	60.7	110000	4	AAI99682	Aai99682	DNA encod
134	16.4	60.7	319608	5	AAH51601	Aah51601	Human chr
135	16.4	60.7	319608	5	AAH51601	Aah51601	Human chr
136	16.2	60.0	435	6	ABN23757	Abn23757	Human chr
137	16.2	60.0	1156	6	ABQ68844	Abq68844	Listeria
138	16.2	60.0	1553	6	ABQ71204	Abq71204	Listeria
139	16.2	60.0	1605	9	AAK685595	Aak685595	DNA encod
140	16.2	60.0	1605	9	AAK685595	Aak685595	DNA encod
141	16.2	60.0	2239	5	AAK685595	Aak685595	DNA encod
142	16.2	60.0	2784	6	ABQ70687	Abq70687	Listeria
143	16.2	60.0	3128	9	ADK52465	Adk52465	Primary r
144	16.2	60.0	5821	2	AAK586686	Aak586686	DNA encod
145	16.2	60.0	8251	4	AAK64961	Aak64961	Human chr
146	16.2	60.0	8251	4	AAK64961	Aak64961	Human chr
147	16.2	60.0	12905	3	AAZ99101	Aaz99101	S. fradia
148	16.2	60.0	12905	3	AAZ99101	Aaz99101	S. fradia
149	16.2	60.0	14070	4	AAK70465	Aak70465	Human chr
150	16.2	60.0	18060	4	ABU14126	Abu14126	Drosophila
151	16.2	60.0	25	8	ACI24457	Act24457	Human mic
152	16.2	60.0	25	8	ACI24457	Act24457	Human mic
153	16.2	60.0	25	8	ACI24457	Act24457	Human mic
154	16.2	60.0	516	5	ABV05086	Abv05086	Human pro
155	16.2	60.0	523	6	ABQ51047	Abq51047	Oligonuc
156	16.2	60.0	523	6	ABQ51046	Abq51046	Oligonuc
157	16.2	60.0	576	6	ABQ91695	Abq91695	M. capsul
158	16.2	60.0	729	7	ACA30336	Act30336	Prokaryot
159	16.2	60.0	866	2	AAK79681	Aak79681	Cynomolg
160	16.2	60.0	866	2	AAK79681	Aak79681	Cynomolg
161	16.2	60.0	866	2	AAK79681	Aak79681	Cynomolg
162	16.2	60.0	866	2	AAK79681	Aak79681	Cynomolg
163	16.2	60.0	866	2	AAK79681	Aak79681	Cynomolg
164	16.2	60.0	866	2	AAK79681	Aak79681	Cynomolg
165	16.2	60.0	866	2	AAK79681	Aak79681	Cynomolg
166	16.2	60.0	866	2	AAK79681	Aak79681	Cynomolg
167	16.2	60.0	866	2	AAK79681	Aak79681	Cynomolg
168	16.2	60.0	866	2	AAK79681	Aak79681	Cynomolg
169	16.2	60.0	866	2	AAK79681	Aak79681	Cynomolg

```
243 15.8 58.5 3831 5 AAS71727 DNA encod
244 15.8 58.5 3949 6 ABK89101 Aspergill
245 15.8 58.5 4312 7 ABZ42895 Human met
246 15.8 58.5 4519 7 ABT43574 Human cla
247 15.8 58.5 4539 9 ADC17324
248 15.8 58.5 5883 3 AAA88981 Toxoplasm
249 15.8 58.5 6560 4 ABL05125 Drosophil
250 15.8 58.5 6774 4 ABL20811 Drosophil
251 15.8 58.5 8411 5 ABA18086 Human ner
252 15.8 58.5 12017 7 ABS58302 Novel hum
253 15.8 58.5 13438 4 ABL20810 Drosophil
254 15.8 58.5 16092 4 ABL05124 Drosophil
255 15.8 58.5 21565 8 ADA02555 Mouse Ly6
256 15.8 58.5 21565 9 ADB72293 Mouse Ly6
257 15.8 58.5 24652 8 ADA02513
258 15.8 58.5 24652 9 ADB72251 Mouse Ras
259 15.8 58.5 34450 8 ACF04005 Human CRF
260 15.8 58.5 34757 8 ACF04006 Human CRF
261 15.8 58.5 38918 9 ADC87242
262 15.8 58.5 89047 4 AAF28547 Genomic f
263 15.8 58.5 90583 7 ACD13447 Human DNA
264 15.8 58.5 110000 2 AAX91990_01 Continuation (2 of
265 15.8 58.5 110000 6 AQC67196_0 Listeria
266 15.8 58.5 110000 6 AQC69245_24 Continuation (26 o
267 15.8 58.5 110000 6 AQC69245_25 Continuation (25 o
268 15.6 57.8 28 6 ABK66062 Human gen
269 15.6 57.8 100 7 ACD76997 E. coli K
270 15.6 57.8 256 6 ABL76576 Corn tass
271 15.6 57.8 270 7 ABX82202 Corn ear-
272 15.6 57.8 313 7 ABX85407 Corn ear-
273 15.6 57.8 347 3 ABA30900 Breast ca
274 15.6 57.8 360 6 ABV78204 Human ID3
275 15.6 57.8 360 6 ABZ35780 Human ID3
276 15.6 57.8 360 6 ABX10023 Human ID3
277 15.6 57.8 360 6 ABL91745 Human pol
278 15.6 57.8 429 8 ACH21548 Human adu
279 15.6 57.8 458 4 AAL82306 Human pol
280 15.6 57.8 465 8 ACH44238 Human foe
281 15.6 57.8 495 4 AAI22113 Probe #12
282 15.6 57.8 495 4 ABA67192 Human foe
283 15.6 57.8 495 4 AAI47408 Probe #16
284 15.6 57.8 495 4 ABA49278 Human bre
285 15.6 57.8 495 4 ABA34287 Probe #12
286 15.6 57.8 495 4 AKA43370 Human bon
287 15.6 57.8 495 4 AAK15636 Human bra
288 15.6 57.8 495 4 ABS40961 Human liv
289 15.6 57.8 495 5 AAI07811 Probe #78
290 15.6 57.8 495 6 ABS15375 Human gen
291 15.6 57.8 687 3 AAF13257 Aspergill
292 15.6 57.8 756 2 AQC20191 Ascorbic
293 15.6 57.8 840 6 ABN67629 Streptoco
294 15.6 57.8 843 7 ACA50767 Prokaryot
295 15.6 57.8 982 2 AQA44245 HEIR-1 ge
296 15.6 57.8 982 6 AQC88114 Human ost
297 15.6 57.8 982 7 ACC46763 Human COP
298 15.6 57.8 1020 7 ACC42715 Epoxide h
299 15.6 57.8 1194 4 AAF88323 S. spinos
300 15.6 57.8 1203 6 AQC88115 Human ost
```

## ALIGNMENTS

```
RESULT 1
AAI99682_25/c
Continuation (26 of 45) of AAI99682 from base 2500001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682
WP Fragment Name Begin End
WP AAI99682_00 1 110000
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000
WP AAI99682_09 900001 1010000
WP AAI99682_10 1000001 1110000
WP AAI99682_11 1100001 1210000
WP AAI99682_12 1200001 1310000
WP AAI99682_13 1300001 1410000
WP AAI99682_14 1400001 1510000
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000
WP AAI99682_17 1700001 1810000
```

```
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000
WP AAI99682_09 900001 1010000
WP AAI99682_10 1000001 1110000
WP AAI99682_11 1100001 1210000
WP AAI99682_12 1200001 1310000
WP AAI99682_13 1300001 1410000
WP AAI99682_14 1400001 1510000
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000
WP AAI99682_17 1700001 1810000
WP AAI99682_18 1800001 1910000
WP AAI99682_19 1900001 2010000
WP AAI99682_20 2000001 2110000
WP AAI99682_21 2100001 2210000
WP AAI99682_22 2200001 2310000
WP AAI99682_23 2300001 2410000
WP AAI99682_24 2400001 2510000
WP AAI99682_25 2500001 2610000
WP AAI99682_26 2600001 2710000
WP AAI99682_27 2700001 2810000
WP AAI99682_28 2800001 2910000
WP AAI99682_29 2900001 3010000
WP AAI99682_30 3000001 3110000
WP AAI99682_31 3100001 3210000
WP AAI99682_32 3200001 3310000
WP AAI99682_33 3300001 3410000
WP AAI99682_34 3400001 3510000
WP AAI99682_35 3500001 3610000
WP AAI99682_36 3600001 3710000
WP AAI99682_37 3700001 3810000
WP AAI99682_38 3800001 3910000
WP AAI99682_39 3900001 4010000
WP AAI99682_40 4000001 4110000
WP AAI99682_41 4100001 4210000
WP AAI99682_42 4200001 4310000
WP AAI99682_43 4300001 4410000
WP AAI99682_44 4400001 4411529
```

Query Match 100.0%; Score 27; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.0038;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Oy 1 CACCCGCGCTTTAGGATCGACACCTGA 27
|||||
Db 104297 CACCCGCGCTTTAGGATCGACACCTGA 104271
```

## RESULT 2

```
AAI99682_26/c
Continuation (27 of 45) of AAI99682 from base 2600001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682
WP Fragment Name Begin End
WP AAI99682_00 1 110000
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000
WP AAI99682_09 900001 1010000
WP AAI99682_10 1000001 1110000
WP AAI99682_11 1100001 1210000
WP AAI99682_12 1200001 1310000
WP AAI99682_13 1300001 1410000
WP AAI99682_14 1400001 1510000
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000
WP AAI99682_17 1700001 1810000
```

WP AAI99682\_18 1800001 1910000  
WP AAI99682\_19 1900001 2010000  
WP AAI99682\_20 2000001 2110000  
WP AAI99682\_21 2100001 2210000  
WP AAI99682\_22 2200001 2310000  
WP AAI99682\_23 2300001 2410000  
WP AAI99682\_24 2400001 2510000  
WP AAI99682\_25 2500001 2610000  
WP AAI99682\_26 2600001 2710000  
WP AAI99682\_27 2700001 2810000  
WP AAI99682\_28 2800001 2910000  
WP AAI99682\_29 2900001 3010000  
WP AAI99682\_30 3000001 3110000  
WP AAI99682\_31 3100001 3210000  
WP AAI99682\_32 3200001 3310000  
WP AAI99682\_33 3300001 3410000  
WP AAI99682\_34 3400001 3510000  
WP AAI99682\_35 3500001 3610000  
WP AAI99682\_36 3600001 3710000  
WP AAI99682\_37 3700001 3810000  
WP AAI99682\_38 3800001 3910000  
WP AAI99682\_39 3900001 4010000  
WP AAI99682\_40 4000001 4110000  
WP AAI99682\_41 4100001 4210000  
WP AAI99682\_42 4200001 4310000  
WP AAI99682\_43 4300001 4410000  
WP AAI99682\_44 4400001 4411529

Query Match 100.0%; Score 27; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.0038;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27  
DB 4297 CACCCGCGCTTTAGGATCGACACCTGA 4271

RESULT 3 25/c  
AAI99683\_25/c  
Continuation (26 of 44) of AAI99683 from base 2500001 (Mycobacterium tuberculosis strain  
WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683  
WP Fragment Name Begin End  
WP AAI99683\_01 1000001 1100000  
WP AAI99683\_02 200001 210000  
WP AAI99683\_03 300001 310000  
WP AAI99683\_04 400001 410000  
WP AAI99683\_05 500001 510000  
WP AAI99683\_06 600001 610000  
WP AAI99683\_07 700001 710000  
WP AAI99683\_08 800001 810000  
WP AAI99683\_09 900001 910000  
WP AAI99683\_10 1000001 110000  
WP AAI99683\_11 1100001 1210000  
WP AAI99683\_12 1200001 1310000  
WP AAI99683\_13 1300001 1410000  
WP AAI99683\_14 1400001 1510000  
WP AAI99683\_15 1500001 1610000  
WP AAI99683\_16 1600001 1710000  
WP AAI99683\_17 1700001 1810000  
WP AAI99683\_18 1800001 1910000  
WP AAI99683\_19 1900001 2010000  
WP AAI99683\_20 2000001 210000  
WP AAI99683\_21 2100001 2210000  
WP AAI99683\_22 2200001 2310000  
WP AAI99683\_23 2300001 2410000  
WP AAI99683\_24 2400001 2510000  
WP AAI99683\_25 2500001 2610000  
WP AAI99683\_26 2600001 2710000  
WP AAI99683\_27 2700001 2810000  
WP AAI99683\_28 2800001 2910000  
WP AAI99683\_29 2900001 3010000  
WP AAI99683\_30 3000001 3110000

WP AAI99683\_31 3100001 3210000  
WP AAI99683\_32 3200001 3310000  
WP AAI99683\_33 3300001 3410000  
WP AAI99683\_34 3400001 3510000  
WP AAI99683\_35 3500001 3610000  
WP AAI99683\_36 3600001 3710000  
WP AAI99683\_37 3700001 3810000  
WP AAI99683\_38 3800001 3910000  
WP AAI99683\_39 3900001 4010000  
WP AAI99683\_40 4000001 4110000  
WP AAI99683\_41 4100001 4210000  
WP AAI99683\_42 4200001 4310000  
WP AAI99683\_43 4300001 4403765

Query Match 100.0%; Score 27; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.0038;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27  
DB 100107 CACCCGCGCTTTAGGATCGACACCTGA 100081

RESULT 4 26/c  
AAI99683\_26/c  
Continuation (27 of 44) of AAI99683 from base 2600001 (Mycobacterium tuberculosis strain  
WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683  
WP Fragment Name Begin End  
WP AAI99683\_00 1000001 110000  
WP AAI99683\_01 1000001 210000  
WP AAI99683\_02 200001 310000  
WP AAI99683\_03 300001 410000  
WP AAI99683\_04 400001 510000  
WP AAI99683\_05 500001 610000  
WP AAI99683\_06 600001 710000  
WP AAI99683\_07 700001 810000  
WP AAI99683\_08 800001 910000  
WP AAI99683\_09 900001 1010000  
WP AAI99683\_10 1000001 1110000  
WP AAI99683\_11 1100001 1210000  
WP AAI99683\_12 1200001 1310000  
WP AAI99683\_13 1300001 1410000  
WP AAI99683\_14 1400001 1510000  
WP AAI99683\_15 1500001 1610000  
WP AAI99683\_16 1600001 1710000  
WP AAI99683\_17 1700001 1810000  
WP AAI99683\_18 1800001 1910000  
WP AAI99683\_19 1900001 2010000  
WP AAI99683\_20 2000001 2110000  
WP AAI99683\_21 2100001 2210000  
WP AAI99683\_22 2200001 2310000  
WP AAI99683\_23 2300001 2410000  
WP AAI99683\_24 2400001 2510000  
WP AAI99683\_25 2500001 2610000  
WP AAI99683\_26 2600001 2710000  
WP AAI99683\_27 2700001 2810000  
WP AAI99683\_28 2800001 2910000  
WP AAI99683\_29 2900001 3010000  
WP AAI99683\_30 3000001 3110000  
WP AAI99683\_31 3100001 3210000  
WP AAI99683\_32 3200001 3310000  
WP AAI99683\_33 3300001 3410000  
WP AAI99683\_34 3400001 3510000  
WP AAI99683\_35 3500001 3610000  
WP AAI99683\_36 3600001 3710000  
WP AAI99683\_37 3700001 3810000  
WP AAI99683\_38 3800001 3910000  
WP AAI99683\_39 3900001 4010000  
WP AAI99683\_40 4000001 4110000  
WP AAI99683\_41 4100001 4210000  
WP AAI99683\_42 4200001 4310000  
WP AAI99683\_43 4300001 4403765

Query Match 100.0%; Score 27; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.0038;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACTGA 27  
DB 107 CACCGCGCTTTAGGATCGACACTGA 81

RESULT 5  
ID ABZ16792 standard; DNA; 2000 BP.  
XX AC ABZ16792;  
XX DT 21-JAN-2003 (first entry)  
XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 4597.  
XX KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.  
XX OS Arabidopsis thaliana.  
XX PN WO200216655-A2.  
XX PD 28-FEB-2002.  
XX PF 24-AUG-2001; 2001WO-US026685.  
XX PR 24-AUG-2000; 2000US-0227866P.  
XX PR 26-JAN-2001; 2001US-0264647P.  
XX PR 22-JUN-2001; 2001US-0300111P.  
XX PA (SCRI) SCRIPPS RES INST.  
XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.  
XX PI Harper JF, Kreps J, Wang X, Zhu T;  
XX WPI; 2002-304127/34.  
XX PT Identifying a stress condition to which a plant cell has been exposed and  
XX PT producing plants with increased tolerance to these abiotic stresses.  
XX PS Claim 144; SEQ ID NO 4597; 577pp + Sequence Listing; English.  
XX CC The invention relates to identifying a stress condition to which a plant  
XX CC cell has been exposed, comprising: (a) contacting nucleic acid  
XX CC representative of expressed polynucleotides in the plant cell with an  
XX CC array or probes representative of the plant cell genome; and (b)  
XX CC detecting a profile of expressed polynucleotides in the plant cell  
XX CC characteristic of a stress response. The method is useful in the  
XX CC production of transgenic plants, cells and seeds and in producing plants  
XX CC with increased tolerance to abiotic stress. The present sequence is that  
XX CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used  
XX CC in methods of the invention. Note: The sequence data for this patent is  
XX CC not represented in the printed specification but is based on sequence  
XX CC information supplied to Derwent by the European Patent Office  
XX SQ Sequence 2000 BP; 634 A; 351 C; 308 G; 707 T; 0 U; 0 Other;

Query Match 68.9%; Score 18.6; DB 6; Length 2000;  
Best Local Similarity 84.0%; Pred. No. 36;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACTG 26  
DB 783 ACCCGCGCTTTAGGATCGACACTG 807

RESULT 6  
ID AAH67093/c  
XX ID AAH67093 standard; DNA; 1287 BP.  
XX

AAH67093;  
26-SEP-2001 (first entry)  
C glutamicum coding sequence fragment SEQ ID NO: 2128.  
Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
organic acid synthesis; ds.  
Corynebacterium glutamicum.  
EPI108790-A2.  
20-JUN-2001.  
18-DEC-2000; 2000EP-00127688.  
16-DEC-1999; 99JP-00377484.  
07-APR-2000; 2000JP-00159162.  
03-AUG-2000; 2000JP-00280988.  
(KYOW) KYOWA HAKKO KOGYO KK.  
Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
Tateishi N, Senoh A, Ikeda M, Ozaki A;  
WPI; 2001-376931/40.  
P-PSDE; AAG91874.  
Novel polynucleotides derived from Coryneform bacteria, for identifying  
mutation point of a gene, measuring expression of a gene, analyzing  
expression profile or pattern of a gene and identifying homologous gene.  
Claim 8; SEQ ID NO 2128; 246pp + Sequence Listing; English.  
The present invention provides a number of nucleotide and protein  
sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
are useful for identifying the mutation point of a gene derived from a  
mutant of coryneform bacterium, measuring expression amount and analysing  
the expression profile or expression pattern of a gene derived from  
Coryneform bacterium, and identifying a homologue of a gene derived from  
Coryneform bacterium. Coryneform bacteria are useful for producing amino  
acids, nucleic acids, vitamins, saccharides and organic acids,  
particularly L-lysine. The present sequence is a nucleic acid described  
in the exemplification of the invention. Note: The sequence data for this  
patent did not form part of the printed specification, but was obtained  
in electronic format directly from the European Patent Office  
SQ Sequence 1287 BP; 258 A; 392 C; 338 G; 299 T; 0 U; 0 Other;

Query Match 66.7%; Score 18; DB 5; Length 1287;  
Best Local Similarity 80.8%; Pred. No. 68;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACTG 26  
DB 90 CACCGCGCTTTAGGATCGACACTG 65

RESULT 7  
ID AAF68053/c  
XX ID AAF68053 standard; DNA; 1335 BP.  
XX AC AAF68053;  
XX DT 11-APR-2001 (first entry)  
XX DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:621.  
XX KW Corynebacterium glutamicum; brevbacterium lactofermentum; MCT;  
KW membrane construction and membrane transport protein; petroleum spill;  
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;  
KW identification; microorganism; fine chemical production; transformation;

genome mapping; genetic engineering; ds.

Corynebacterium glutamicum.

WO200100805-A2.

04-JAN-2001.

23-JUN-2000; 2000WO-IB000926.

25-JUN-1999; 99US-0141031P.

08-JUL-1999; 99DE-01031454.

08-JUL-1999; 99DE-01031478.

08-JUL-1999; 99DE-01031563.

09-JUL-1999; 99DE-01032122.

09-JUL-1999; 99DE-01032124.

09-JUL-1999; 99DE-01032125.

09-JUL-1999; 99DE-01032128.

09-JUL-1999; 99DE-01032180.

09-JUL-1999; 99DE-01032182.

09-JUL-1999; 99DE-01032190.

09-JUL-1999; 99DE-01032193.

09-JUL-1999; 99DE-01032209.

09-JUL-1999; 99DE-01032212.

09-JUL-1999; 99DE-01032227.

09-JUL-1999; 99DE-01032228.

09-JUL-1999; 99DE-01032229.

09-JUL-1999; 99DE-01032230.

14-JUL-1999; 99DE-01032237.

14-JUL-1999; 99DE-01033005.

27-AUG-1999; 99DE-01040764.

27-AUG-1999; 99DE-01040765.

27-AUG-1999; 99DE-01040766.

27-AUG-1999; 99DE-01040830.

27-AUG-1999; 99DE-01040831.

27-AUG-1999; 99DE-01040832.

27-AUG-1999; 99DE-01040833.

31-AUG-1999; 99DE-01041378.

31-AUG-1999; 99DE-01041379.

31-AUG-1999; 99DE-01041395.

03-SEP-1999; 99DE-01042077.

03-SEP-1999; 99DE-01042078.

03-SEP-1999; 99DE-01042079.

03-SEP-1999; 99DE-01042088.

(BADI ) BASF AG.

Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

WPI; 2001-071486/08.

P-PSDB; AAB76820.

Corynebacterium glutamicum nucleic acids encoding membrane construction and membrane transport proteins or their portions, useful for typing or identifying C. glutamicum or related bacteria, and as markers for transformation.

Claim 3; Page 1040-1040; 1119pp; English.

AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane construction and membrane transport (MCT) proteins given in AAB76510 to AAB76847. The MCT nucleic acids and proteins are useful in the identification of microorganisms which can be used to produce fine chemicals, for modulating fine chemical production in C. glutamicum or related bacteria (e.g. Brevibacterium lactofermentum), the typing or identification of C. glutamicum or related bacteria, as reference points for mapping C. glutamicum genome, and as markers for transformation. AAF68082 and AAF68082 represent sequencing primers which are used in an example from the present invention

Sequence 1335 BP; 268 A; 411 C; 344 G; 312 T; 0 U; 0 Other;

Query Match 66.7%; Score 18; DB 4; Length 1335;

Best Local Similarity 80.8%; Pred. No. 68;

Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCTG 26

Db 190 CAGCGCTCTTTAGGAGCGACACTG 165

RESULT 8

AAF68052/c

ID AAF68052 standard; DNA; 1410 BP.

XX AC AAF68052;

XX DT 11-APR-2001 (first entry)

XX DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:619.

XX KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT; membrane construction and membrane transport protein; petroleum spill; hydrocarbon degradation; gram positive aerobic bacterium; marker; identification; microorganism; fine chemical production; transformation; genome mapping; genetic engineering; ds.

XX OS Corynebacterium glutamicum.

XX PN WO200100805-A2.

XX PD 04-JAN-2001.

XX PF 23-JUN-2000; 2000WO-IB000926.

XX PR 25-JUN-1999; 99US-0141031P.

XX PR 08-JUL-1999; 99DE-01031454.

XX PR 08-JUL-1999; 99DE-01031478.

XX PR 09-JUL-1999; 99DE-01031563.

XX PR 09-JUL-1999; 99DE-01032122.

XX PR 09-JUL-1999; 99DE-01032124.

XX PR 09-JUL-1999; 99DE-01032125.

XX PR 09-JUL-1999; 99DE-01032128.

XX PR 09-JUL-1999; 99DE-01032180.

XX PR 09-JUL-1999; 99DE-01032182.

XX PR 09-JUL-1999; 99DE-01032190.

XX PR 09-JUL-1999; 99DE-01032209.

XX PR 09-JUL-1999; 99DE-01032212.

XX PR 09-JUL-1999; 99DE-01032227.

XX PR 09-JUL-1999; 99DE-01032228.

XX PR 09-JUL-1999; 99DE-01032229.

XX PR 14-JUL-1999; 99DE-01032230.

XX PR 14-JUL-1999; 99DE-01032237.

XX PR 14-JUL-1999; 99DE-01033006.

XX PR 27-AUG-1999; 99DE-01040764.

XX PR 27-AUG-1999; 99DE-01040765.

XX PR 27-AUG-1999; 99DE-01040766.

XX PR 27-AUG-1999; 99DE-01040830.

XX PR 27-AUG-1999; 99DE-01040831.

XX PR 27-AUG-1999; 99DE-01040832.

XX PR 27-AUG-1999; 99DE-01040833.

XX PR 31-AUG-1999; 99DE-01041378.

XX PR 31-AUG-1999; 99DE-01041379.

XX PR 31-AUG-1999; 99DE-01041395.

XX PR 03-SEP-1999; 99DE-01042077.

XX PR 03-SEP-1999; 99DE-01042078.

XX PR 03-SEP-1999; 99DE-01042079.

XX PR 03-SEP-1999; 99DE-01042088.

XX PA (BADI ) BASF AG.

XX PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

XX PT WPI; 2001-071486/08.

XX DR P-PSDB; AAB76820.

XX XX Corynebacterium glutamicum nucleic acids encoding membrane construction and membrane transport proteins or their portions, useful for typing or identifying C. glutamicum or related bacteria, and as markers for transformation.

XX XX Claim 3; Page 1040-1040; 1119pp; English.

XX XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane construction and membrane transport (MCT) proteins given in AAB76510 to AAB76847. The MCT nucleic acids and proteins are useful in the identification of microorganisms which can be used to produce fine chemicals, for modulating fine chemical production in C. glutamicum or related bacteria (e.g. Brevibacterium lactofermentum), the typing or identification of C. glutamicum or related bacteria, as reference points for mapping C. glutamicum genome, and as markers for transformation. AAF68082 and AAF68082 represent sequencing primers which are used in an example from the present invention

XX XX Sequence 1335 BP; 268 A; 411 C; 344 G; 312 T; 0 U; 0 Other;





CC probes in diagnostic assays to detect and quantitate the presence of  
 CC similar nucleic acid sequences in samples, and therefore which patients  
 CC may be in need of restorative therapy. The proteins may also be used as  
 CC antigens in the production of antibodies against bone marrow proteins and  
 CC in assays to identify modulators of their expression and activity. The  
 CC anti-bone marrow protein antibodies and antagonists may also be used to  
 CC down regulate expression and activity. The antibodies may also be used as  
 CC diagnostic agents for detecting the presence of the protein in samples  
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins may be  
 CC used to regulate haematopoiesis activity, and consequently in the  
 CC treatment of myeloid or lymph cell disorders; in tissue regeneration,  
 CC such as wound healing; as a nutritional supplement; and in treatment of  
 CC immune disorders such as severe combined immunodeficiency (SCID)

XX  
 SQ Sequence 766 BP; 186 A; 184 C; 148 G; 248 T; 0 U; 0 Other;

Query Match 65.2%; Score 17.6; DB 4; Length 766;  
 Best Local Similarity 83.3%; Pred. No. 1e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 CACCCCGCGCTTAGGATCGACACC 24  
 Db 658 CACCCCGCGCTTAGGATCGACACC 681

RESULT 11  
 AAS23029/C  
 ID AAS23029 standard; DNA; 771 BP.  
 XX AAS23029;  
 AC  
 AC  
 DT 24-OCT-2001 (first entry)  
 XX  
 DE DNA encoding novel bone marrow polypeptide #123.  
 KW Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;  
 KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;  
 KW wound healing; nutritional supplement; immune disorder;  
 KW severe combined immunodeficiency; SCID; ds.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200157187-A2.  
 PN  
 XX 09-AUG-2001.  
 PD  
 XX 05-FEB-2001; 2001WO-US003782.  
 PF  
 XX 03-FEB-2000; 2000US-00496914.  
 PR 20-JUN-2000; 2000US-00598075.  
 PR 19-JUL-2000; 2000US-00620325.  
 PR 30-NOV-2000; 2000US-0250683P.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;  
 PI Ren F, Drmanac RT;  
 PI  
 XX WPI; 2001-488875/53.  
 DR P-PSDB; AAU14724.  
 DR  
 XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic and  
 PT gene therapy.  
 PT  
 XX  
 XX Claim 1; Page 271; 392pp; English.  
 PS  
 XX AAS22907-AAS23099 represent nucleic acids encoding novel bone marrow  
 CC polypeptides. The nucleic acids and corresponding proteins may be used in  
 CC the prevention, diagnosis and treatment of diseases associated with  
 CC inappropriate bone marrow polypeptide expression. For example, to treat  
 CC disorders associated with decreased expression by rectifying mutations or  
 CC deletions in a patient's genome that affect the activity of the  
 CC polypeptides by expressing inactive proteins or to supplement the

CC patient's own production of the polypeptide. Additionally, the nucleic  
 CC acids may be used to produce the polypeptides, by inserting the nucleic  
 CC acids into a host cell and culturing the cell to express the protein. The  
 CC nucleic acid and its complementary sequences may also be used as DNA  
 CC probes in diagnostic assays to detect and quantitate the presence of  
 CC similar nucleic acid sequences in samples, and therefore which patients  
 CC may be in need of restorative therapy. The proteins may also be used as  
 CC antigens in the production of antibodies against bone marrow proteins and  
 CC in assays to identify modulators of their expression and activity. The  
 CC anti-bone marrow protein antibodies and antagonists may also be used to  
 CC down regulate expression and activity. The antibodies may also be used as  
 CC diagnostic agents for detecting the presence of the protein in samples  
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins may be  
 CC used to regulate haematopoiesis activity, and consequently in the  
 CC treatment of myeloid or lymph cell disorders; in tissue regeneration,  
 CC such as wound healing; as a nutritional supplement; and in treatment of  
 CC immune disorders such as severe combined immunodeficiency (SCID)

XX  
 SQ Sequence 771 BP; 132 A; 192 C; 270 G; 177 T; 0 U; 0 Other;

Query Match 65.2%; Score 17.6; DB 4; Length 771;  
 Best Local Similarity 83.3%; Pred. No. 1e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 CACCCCGCGCTTAGGATCGACACC 24  
 Db 109 CACCCCGCGCTTAGGATCGACACC 86

RESULT 12  
 ABS56529  
 ID ABS56529 standard; DNA; 894 BP.  
 XX ABS56529;  
 AC  
 AC  
 DT 27-JAN-2003 (first entry)  
 XX  
 DE C. violaceum DNA encoding Phenylalanine ammonia-lyase, PAL.  
 KW ds; gene; phenylalanine ammonia-lyase; PAL; phenylalanine hydroxylase;  
 KW PAH; para-hydroxycinnamic acid; tyrosine ammonium lyase; TAL; tyrosine;  
 KW cinnamic acid; PHCA; liquid crystal polymer.  
 XX  
 OS Chromobacterium violaceum.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 1..891  
 FT /\*tag= a  
 FT /product= "phenylalanine ammonia-lyase"  
 XX  
 XX WO200290523-A2.  
 PN  
 XX 14-NOV-2002.  
 PD  
 XX 03-MAY-2002; 2002WO-US018551.  
 PF  
 XX 04-MAY-2001; 2001US-0288701P.  
 PR  
 XX (DUPO) DU FONT DE NEMOURS & CO E I.  
 PA  
 XX Qi WW, Sariaslani FS, Tang X;  
 PI  
 XX WPI; 2003-058928/05.  
 DR P-PSDB; ABG71947.  
 DR  
 XX Novel recombinant host for producing para-hydroxycinnamic acid and  
 PT tyrosine, comprises gene encoding tyrosine ammonium lyase activity and  
 PT gene encoding phenylalanine hydroxylase activity.  
 XX  
 XX Example 1; Page 45; 69pp; English.  
 PS  
 XX The invention relates to a recombinant host comprising at least one gene  
 CC encoding a tyrosine ammonium lyase (TAL) activity, and at least one gene

CC encoding a phenylalanine hydroxylase (PAH) activity. Also included are:  
 CC (1) a method for producing para-hydroxycinnamic acid involving: (1)  
 CC providing the recombinant host; and (2) growing the recombinant organism  
 CC in the presence of a fermentable carbon substrate and thus producing para  
 CC -hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)  
 CC providing a recombinant organism comprising at least one gene encoding a  
 CC phenylalanine hydroxylase activity; and (2) growing the recombinant  
 CC organism in the presence of a fermentable carbon substrate to produce  
 CC tyrosine. The recombinant host cell is used for producing para,  
 CC hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the  
 CC production of para-hydroxycinnamic acid (PHCA) which is useful as a  
 CC monomer for the production of liquid crystal polymer. The present  
 CC sequence is C. violaceum DNA encoding Phenylalanine ammonia-lyase, PAH,  
 CC an enzyme which catalyses the conversion of phenylalanine into cinnamic  
 CC acid  
 CC  
 XX  
 SQ Sequence 894 BP; 167 A; 301 C; 274 G; 152 T; 0 U; 0 Other;  
 Query Match 65.2%; Score 17.6; DB 7; Length 894;  
 Best Local Similarity 83.3%; Pred. No. 1e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 ACCCGCGCTTTAGGTCACACCT 25  
 |||||  
 DB 686 ACACGCGCTACCGGATCGACACCT 709  
 |||||  
 RESULT 13  
 ACA25099/C  
 ID ACA25099 standard; DNA; 933 BP.  
 XX  
 AC ACA25099;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Prokaryotic essential gene #6756.  
 XX  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.  
 XX  
 OS Burkholderia fungorum.  
 XX  
 PN WO20027183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR P-PSDB; ABU21229.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 14; SEQ ID NO 12969; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 933 BP; 164 A; 293 C; 321 G; 155 T; 0 U; 0 Other;

Query Match 65.2%; Score 17.6; DB 7; Length 933;  
 Best Local Similarity 83.3%; Pred. No. 1e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CCGCGCGCTTTAGGATCGACACCTG 26  
 |||||  
 DB 486 CTCGAGCTTCAGGATCGTCACTG 463  
 |||||

## RESULT 14

AAS73721  
 ID AAS73721 standard; cDNA; 1195 BP.

AC AAS73721;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #9525.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG09534.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 XX PT responsible for genetic disorders or other traits and to assess  
 XX PT biodiversity.

XX Claim 1; SEQ ID NO 9525; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and polypeptide (II)

XX sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed

CC genes. (II) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is

CC useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic

CC coding sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences

XX

XX SQ Sequence 1195 BP; 208 A; 373 C; 372 G; 241 T; 0 U; 1 Other;

Query Match 65.2%; Score 17.6; DB 5; Length 1195;

Best Local Similarity 83.3%; Pred. No. 1.1e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACC 24

Db 769 CCCCCGCGCTTTATGACCGACCCC 792

RESULT 15

ABN59815

ID ABN59815 standard; cDNA; 4919 BP.

XX

XX AC ABN59815;

XX

XX DT 28-JUN-2002 (first entry)

XX

XX DE Novel human coding sequence SEQ ID NO: 226.

XX

XX KW Human; antianaemic; vulnary; antiinflammatory; immunomodulator;

XX KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;

XX KW neuroprotective; antiparkinsonian; protein therapy; EST;

XX KW expressed sequence tag; gene; ss.

XX

XX OS Homo sapiens.

XX

XX PN WO200222660-A2.

XX

XX PD 21-MAR-2002.

XX

XX PF 10-SEP-2001; 2001WO-US026015.

XX

XX PR 11-SEP-2000; 2000US-00659671.

XX

XX PA (HYSE-) HYSEQ INC.

XX

XX PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

XX PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX

XX WPI; 2002-292408/33.

DR P-PSDB; ABB97402.

XX

XX An isolated polynucleotide for treating diseases associated with its

PT encoded polypeptide such as cancer and multiple sclerosis.

XX

XX Claim 1; SEQ ID NO 226; 509pp; English.

XX

CC The present invention provides the protein and coding sequences of 444

CC novel human proteins. These were isolated from expressed sequences tags

CC (ESTs). They can be used to stimulate cell growth, to regulate

CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth

CC e.g. in burn treatment, to regulate the immune system e.g. to treat

CC multiple sclerosis, to regulate activin or inhibin e.g. to treat

CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke

CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.

CC rheumatoid arthritis, and to treat nervous system disorders e.g.

CC Parkinson's disease. The present sequence is a coding sequence of the

CC invention

XX SQ Sequence 4919 BP; 1366 A; 1264 C; 1177 G; 1112 T; 0 U; 0 Other;

Query Match 65.2%; Score 17.6; DB 6; Length 4919;

Best Local Similarity 83.3%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACC 24

Db 1446 CACCCCGCGCTTATGACCGACCC 1469

RESULT 16

ACA38547/C

ID ACA38547 standard; DNA; 1789 BP.

XX

XX AC ACA38547;

XX

XX DT 19-JUN-2003 (first entry)

XX

XX DE Prokaryotic essential gene #20204.

XX

XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;

XX KW drug design; gene.

XX

XX OS Mycobacterium bovis.

XX

XX PN WO200277183-A2.

XX

XX PD 03-OCT-2002.

XX

XX PF 21-MAR-2002; 2002WO-US009107.

XX

XX PR 21-MAR-2001; 2001US-00815242.

XX

XX PR 06-SEP-2001; 2001US-00948993.

XX

XX PR 25-OCT-2001; 2001US-0342823P.

XX

XX PR 08-FEB-2002; 2002US-00072851.

XX

XX PR 06-MAR-2002; 2002US-0362699P.

XX

XX PA (ELIT-) ELITRA PHARM INC.

XX

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;

XX

XX DR WPI; 2003-029926/02.

XX

XX DR P-PSDB; ABU34677.

XX

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX for homologous nucleic acids required for cellular proliferation to

XX isolate candidate molecules for rational drug discovery programs.

XX

XX PS Claim 14; SEQ ID NO 26417; 1766pp; English.

XX

XX The invention relates to an isolated nucleic acid comprising any one of

XX the 6213 antisense sequences given in the specification where expression

XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX (1) a vector comprising a promoter operably linked to the nucleic acid

XX encoding a polypeptide whose expression is inhibited by the antisense

XX nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX polypeptide or its fragment whose expression is inhibited by the

XX antisense nucleic acid; (4) an antibody capable of specifically binding

the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 1789 BP; 362 A; 573 C; 586 G; 268 T; 0 U; 0 Other;  
Query Match 64.4%; Score 17.4; DB 7; Length 1789;  
Best Local Similarity 77.8%; Pred. No. 1.4e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCGCGCTTTAGGATCGACACTGA 27  
Db 1293 CGCCGCGCGCAGGATCGTCACTTA 1267

RESULT 17  
AAH52081/c  
ID AAH52081 standard; DNA; 2283 BP.

AC AAH52081;  
XX  
DT 04-SEP-2001 (first entry)  
XX  
DE Mycobacterium tuberculosis potential drug target gene SEQ ID 135.  
XX  
KW Drug target; growth; organism viability; characterisation; ds.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO200135317-A1.  
XX  
PD 17-MAY-2001.  
XX  
PF 13-NOV-2000; 2000WO-US031152.  
XX  
PR 12-NOV-1999; 99US-0165086P.  
PR 12-NOV-1999; 99US-0165124P.  
PR 01-FEB-2000; 2000US-0179531P.  
XX  
PA (REGC) UNIV CALIFORNIA.  
XX  
PI Eisenberg D, Rotstein SH, Marcotte EM;  
XX  
DR WPI; 2001-329193/34.  
DR P-PSDB; AAH52081.

Identifying nucleotide or polypeptide sequence for use as drug target, involves providing algorithm that analyzes a functional relationship between nucleotide or polypeptide sequences, and comparing the sequences.

Disclosure; Page 143-144; 207pp; English.

This invention relates to a method for identifying a nucleotide or

polypeptide sequence that may be a drug target, or essential for growth or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092 represent DNA encoding proteins AAH51096 - AAH51241, Mycobacterium tuberculosis proteins which are potential drug targets. The DNA and protein sequences are used to illustrate the method of the invention. The method involves providing an unknown nucleotide or polypeptide sequences, and comparing it to a number of sequences along with at least one algorithm capable of analysing a functional relationship between nucleotide and polypeptide sequences. The method is useful for characterising the function of nucleic acids and polypeptides that may be useful as a target for a drug or essential for the growth or viability of an organism

Sequence 2283 BP; 452 A; 757 C; 745 G; 329 T; 0 U; 0 Other;  
Query Match 64.4%; Score 17.4; DB 4; Length 2283;  
Best Local Similarity 77.8%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCGCGCTTTAGGATCGACACTGA 27  
Db 1355 CGCCGCGCGCAGGATCGTCACTTA 1329

RESULT 18  
ACA40856/c  
ID ACA40856 standard; DNA; 2283 BP.

AC ACA40856;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Prokaryotic essential gene #22513.  
XX  
KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR P-PSDB; ABU35986.

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 14; SEQ ID NO 28726; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding



Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCG 19  
Db 2764 CACCCGCGCTTTAGGATCG 2782

RESULT 21  
AAI99682\_40  
Continuation (41 of 45) of AAI99682 from base 4000001 (Mycobacterium tuberculosis strain  
WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682  
WP Fragment Name Begin End  
WP AAI99682\_00 1 110000  
WP AAI99682\_01 100001 210000  
WP AAI99682\_02 200001 310000  
WP AAI99682\_03 300001 410000  
WP AAI99682\_04 400001 510000  
WP AAI99682\_05 500001 610000  
WP AAI99682\_06 600001 710000  
WP AAI99682\_07 700001 810000  
WP AAI99682\_08 800001 910000  
WP AAI99682\_09 900001 1010000  
WP AAI99682\_10 1000001 1110000  
WP AAI99682\_11 1100001 1210000  
WP AAI99682\_12 1200001 1310000  
WP AAI99682\_13 1300001 1410000  
WP AAI99682\_14 1400001 1510000  
WP AAI99682\_15 1500001 1610000  
WP AAI99682\_16 1600001 1710000  
WP AAI99682\_17 1700001 1810000  
WP AAI99682\_18 1800001 1910000  
WP AAI99682\_19 1900001 2010000  
WP AAI99682\_20 2000001 2110000  
WP AAI99682\_21 2100001 2210000  
WP AAI99682\_22 2200001 2310000  
WP AAI99682\_23 2300001 2410000  
WP AAI99682\_24 2400001 2510000  
WP AAI99682\_25 2500001 2610000  
WP AAI99682\_26 2600001 2710000  
WP AAI99682\_27 2700001 2810000  
WP AAI99682\_28 2800001 2910000  
WP AAI99682\_29 2900001 3010000  
WP AAI99682\_30 3000001 3110000  
WP AAI99682\_31 3100001 3210000  
WP AAI99682\_32 3200001 3310000  
WP AAI99682\_33 3300001 3410000  
WP AAI99682\_34 3400001 3510000  
WP AAI99682\_35 3500001 3610000  
WP AAI99682\_36 3600001 3710000  
WP AAI99682\_37 3700001 3810000  
WP AAI99682\_38 3800001 3910000  
WP AAI99682\_39 3900001 4010000  
WP AAI99682\_40 4000001 4110000  
WP AAI99682\_41 4100001 4210000  
WP AAI99682\_42 4200001 4310000  
WP AAI99682\_43 4300001 4410000  
WP AAI99682\_44 4400001 4411529

Query Match 64.4%; Score 17.4; DB 4; Length 110000;  
Best Local Similarity 77.8%; Pred. No. 2.5e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCTGA 27  
Db 51526 CGCCCGCGCGCCAGGATCGTCACCTTA 51552

RESULT 22  
AAI99683\_40  
Continuation (41 of 44) of AAI99683 from base 4000001 (Mycobacterium tuberculosis strain  
WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683  
WP Fragment Name Begin End  
WP AAI99683\_00 1 110000

AAI99683\_01 100001 210000  
AAI99683\_02 200001 310000  
AAI99683\_03 300001 410000  
AAI99683\_04 400001 510000  
AAI99683\_05 500001 610000  
AAI99683\_06 600001 710000  
AAI99683\_07 700001 810000  
AAI99683\_08 800001 910000  
AAI99683\_09 900001 1010000  
AAI99683\_10 1000001 1110000  
AAI99683\_11 1100001 1210000  
AAI99683\_12 1200001 1310000  
AAI99683\_13 1300001 1410000  
AAI99683\_14 1400001 1510000  
AAI99683\_15 1500001 1610000  
AAI99683\_16 1600001 1710000  
AAI99683\_17 1700001 1810000  
AAI99683\_18 1800001 1910000  
AAI99683\_19 1900001 2010000  
AAI99683\_20 2000001 2110000  
AAI99683\_21 2100001 2210000  
AAI99683\_22 2200001 2310000  
AAI99683\_23 2300001 2410000  
AAI99683\_24 2400001 2510000  
AAI99683\_25 2500001 2610000  
AAI99683\_26 2600001 2710000  
AAI99683\_27 2700001 2810000  
AAI99683\_28 2800001 2910000  
AAI99683\_29 2900001 3010000  
AAI99683\_30 3000001 3110000  
AAI99683\_31 3100001 3210000  
AAI99683\_32 3200001 3310000  
AAI99683\_33 3300001 3410000  
AAI99683\_34 3400001 3510000  
AAI99683\_35 3500001 3610000  
AAI99683\_36 3600001 3710000  
AAI99683\_37 3700001 3810000  
AAI99683\_38 3800001 3910000  
AAI99683\_39 3900001 4010000  
AAI99683\_40 4000001 4110000  
AAI99683\_41 4100001 4210000  
AAI99683\_42 4200001 4310000  
AAI99683\_43 4300001 4403765

Query Match 64.4%; Score 17.4; DB 4; Length 110000;  
Best Local Similarity 77.8%; Pred. No. 2.5e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCTGA 27  
Db 43594 CGCCCGCGCGCCAGGATCGTCACCTTA 43620

RESULT 23  
ABK74646/c  
ID ABK74646 standard; DNA; 999 BP.  
XX  
AC ABK74646;  
XX  
DT 13-AUG-2002 (first entry)  
XX  
DE Bacillus licheniformis genomic sequence tag (GST) #1937.  
XX  
KW Differential gene expression; genomic sequenced tag; GST;  
KW altered culture condition; environmental stress;  
KW physiological provocation; ds.  
XX  
OS Bacillus licheniformis.  
XX  
PN WO200229113-A2.  
XX  
PD 11-APR-2002.  
XX





KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000668.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
PD WPI; 2001-488900/53.  
XX  
PR Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human bone marrow.  
XX  
PS Example 4; SEQ ID NO 19100; 658pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention  
XX  
SQ Sequence 279 BP; 76 A; 64 C; 67 G; 72 T; 0 U; 0 Other;  
Query Match 63.0%; Score 17; DB 4; Length 279;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 2 ACCCGCGCTTAGGATCGACACCTG 26  
DB 36 ACCCTCTCTTTGGGATCCCACTG 12  
RESULT 26  
ABS44201/c  
ID ABS44201 standard; DNA; 279 BP.  
XX  
AC ABS44201;  
XX  
DT 25-FEB-2003 (first entry)  
XX  
DE Human liver single exon probe, SEQ ID No 19191.  
XX  
KW Human; single exon nucleic acid probe; liver; cirrhosis;  
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;  
KW coronary heart disease; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157273-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000664.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR

PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
PD WPI; 2001-488998/53.  
XX  
PR Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human adult liver.  
XX  
PS Claim 4; SEQ ID NO 19191; 658pp; English.  
XX  
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for  
CC measuring human gene expression in a sample derived from human adult  
CC liver, comprising one of 13109 defined nucleotide sequences given in the  
CC specification (or complements/ fragments). The probe hybridises at high  
CC stringency to a nucleic acid molecule expressed in the human adult liver.  
CC (I) may be used for predicting, measuring and displaying gene expression  
CC in samples derived from human adult liver. The genes identified may be  
CC involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
CC associated with coronary heart disease. ABS25011-ABS51005 represent human  
CC liver single exon nucleic acid probes of the invention. Note: The  
CC sequence information for this patent does not appear in the printed  
CC specification but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 279 BP; 76 A; 64 C; 67 G; 72 T; 0 U; 0 Other;  
Query Match 63.0%; Score 17; DB 4; Length 279;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 2 ACCCGCGCTTAGGATCGACACCTG 26  
DB 36 ACCCTCTCTTTGGGATCCCACTG 12  
RESULT 27  
ABS18780/c  
ID ABS18780 standard; DNA; 279 BP.  
XX  
AC ABS18780;  
XX  
DT 19-AUG-2002 (first entry)  
XX  
DE Human genome-derived single exon probe ORF from lung SEQ ID No 18771.  
XX  
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KW pulmonary histiocytosis; lymphangioma; lymphoma;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease; open reading frame; ORF.  
XX  
OS Homo sapiens.  
XX  
PN WO200186003-A2.  
XX  
PD 15-NOV-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000665.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR

PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2002-114183/15.  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples.  
 PT  
 PS Claim 4; SEQ ID NO 18771; 634pp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of probes  
 CC ; the novel set of probes which hybridise at high stringency to a nucleic  
 CC acid expressed in the human lung; measuring gene expression in a sample  
 CC derived from human lung, comprising (a) contacting the array with a  
 CC collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA; and (b) measuring the label detectably bound to each probe of the  
 CC array; identifying exons in a eukaryotic genome, comprising (a)  
 CC algorithmically predicting at least one exon from genomic sequences of  
 CC the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene expression  
 CC analysis, and for identifying exons in a gene, particularly using human  
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
 CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
 CC present sequence is a single exon probe open reading frame of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ffp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 279 BP; 76 A; 64 C; 67 G; 72 T; 0 U; 0 Other;  
 Query Match 63.0%; Score 17; DB 6; Length 279;  
 Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 OY 2 ACCCGCGCTTTAGGATCGACACTG 26  
 DB 36 ACCTCTCTTTGGATCCACACTG 12  
 RESULT 28  
 ABS69598  
 ID ABS69598 standard; DNA; 349 BP.  
 XX  
 XX ABS69598;  
 XX  
 XX 21-NOV-2002 (first entry)  
 DT

XX  
 DE Novel murine polynucleotide isolated using gene trap technology #661.  
 XX  
 KW Mouse; gene trapped sequence; GTS; functional genomic analysis;  
 KW phase display system; gene chip; temporal gene expression;  
 KW tissue specific gene expression; antisense inhibition; gene targeting;  
 KW development disorder; cell differentiation disorder; aging; cancer;  
 KW autoimmune disease; lupus; inflammatory disorder; skin disorder;  
 KW degenerative disorder; ds.  
 XX  
 OS Mus musculus.  
 XX  
 PN US2002102543-A1.  
 XX  
 PD 01-AUG-2002.  
 XX  
 PF 30-NOV-2000; 2000US-00728445.  
 XX  
 PR 01-DEC-1999; 99US-0168358P.  
 XX  
 PA (FRIE/) FRIEDRICH G.  
 PA (ZAMB/) ZAMBROWICZ B.  
 PA (SAND/) SANDS A T.  
 PI Friedrich G, Zambrowicz B, Sands AT;  
 DR WPI; 2002-690598/74.  
 XX  
 PT Novel murine polynucleotides that individually identify novel genes into  
 PT which a retroviral gene trap vector has integrated, useful in genomic  
 PT analysis and in discovery, development of therapeutic and diagnostic  
 PT agents.  
 XX  
 PS Claim 1; Page 222; 295pp; English.  
 XX  
 CC The invention describes an isolated murine polynucleotide (I) comprising  
 CC a contiguous stretch of at least 60 nucleotides of one of 265-677  
 CC nucleotide 991 OMNIBANK gene trapped sequences (GTSs) (S), given in the  
 CC specification. The novel genes and cells are useful in functional genomic  
 CC analysis and in the discovery and development of new therapeutic and  
 CC diagnostic agents and methods. (I) is useful for identifying the coding  
 CC regions of the murine genome, to isolate cDNAs, genomic clones, or full-  
 CC length genes/polynucleotides or homologues, heterologues, paralogues, or  
 CC orthologues that are capable of hybridising to one or more of the GTSs  
 CC under stringent conditions. (I) can be incorporated into a phage display  
 CC system that can be used to screen for proteins, or other ligands, that  
 CC are capable of binding an amino acid sequence encoded by an  
 CC oligonucleotide or polynucleotide sequence in at least one of the TS  
 CC sequences. (I) is useful in addressable arrays, such as gene chips, to  
 CC identify and characterise temporal and tissue specific gene expression,  
 CC to identify the gene of interest from many sources and for genetic  
 CC manipulations such as antisense inhibition and gene targeting. Decreasing  
 CC the level of expression of (I) and/or down regulating the activity of  
 CC peptides or proteins encoded by (I) is useful for treating development  
 CC and cell differentiation disorders, aging, cancer, autoimmune disease,  
 CC lupus, inflammatory disorders, skin disorders and degenerative disorders.  
 CC This sequence represents a murine cDNA isolated using gene trap  
 CC technology  
 XX  
 SQ Sequence 349 BP; 76 A; 95 C; 97 G; 66 T; 0 U; 15 Other;  
 Query Match 63.0%; Score 17; DB 6; Length 349;  
 Best Local Similarity 80.0%; Pred. No. 1.8e+02;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 OY 1 CACCGCGCTTTAGGATCGACACT 25  
 DB 158 CACCGCGCTTCAGATCGACACT 182  
 RESULT 29  
 AAK31460/c  
 ID AAK31460 standard; DNA; 463 BP.

```
XX AAK31460;
AC
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 6017.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000668.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
PS Example 4; SEQ ID NO 6017; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention
XX
SQ Sequence 463 BP; 108 A; 121 C; 111 G; 123 T; 0 U; 0 Other;
Query Match 63.0%; Score 17; DB 4; Length 463;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 ACCCGCGCTTAGGATCGACACTG 26
DB 262 ACCCTCTCTTTGGGATCCCAACTG 238
RESULT 30
ABS31142/c
ID ABS31142 standard; DNA; 463 BP.
XX
AC ABS31142;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver single exon probe, SEQ ID No 6132.
XX
KW Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
```

```
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000664.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488998/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
PS Claim 1; SEQ ID NO 6132; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-ABS51005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 463 BP; 108 A; 121 C; 111 G; 123 T; 0 U; 0 Other;
Query Match 63.0%; Score 17; DB 4; Length 463;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 ACCCGCGCTTAGGATCGACACTG 26
DB 262 ACCCTCTCTTTGGGATCCCAACTG 238
Search completed: June 20, 2004, 10:16:44
Job time : 177.268 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 1261.71 Seconds  
(without alignments)  
639.034 Million cell updates/sec

Title: US-10-624-714-12  
Perfect score: 27  
Sequence: 1 caccgcgcttagatcgacactga 27

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues  
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 300 summaries

Database :

EST:

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hic:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vri:\*
- 28: gb\_gss1:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.2	74.8	324	10	BB504590
C 2	19.6	72.6	630	13	CA055313 ssalrga51
C 3	19.4	71.9	634	28	AQ259762 nbxb0023D
4	19.2	71.1	1130	13	BQ683442 AGENCOURT

19	70.4	483	28	AZ143037	AZ143037 SP 0041.B
19	70.4	533	29	CE809540	CE809540 tigr-gss
18.6	68.9	237	13	BQ973570	BQ973570 QH13D08.
18.6	68.9	443	10	BF149651	BF149651 uy7e01.y
18.6	68.9	603	12	BQ010761	BQ010761 BU010761
18.6	68.9	662	12	BG500183	BG500183 602547620
18.6	68.9	719	11	CNS08HL2	CNS08HL2 tigr-gss
18.6	68.9	757	29	CE802854	CE802854 tigr-gss
18.6	68.9	922	10	BF789780	BF789780 602105331
18.6	68.9	1131	29	AY403727	AY403727 Mus muscu
18.6	68.9	1342	11	AK006855	AK006855 Mus muscu
18.2	67.4	257	13	BU666579	BU666579 Na_L3_39H
18.2	67.4	510	10	BE029790	BE029790 kp34r02.y
18.2	67.4	542	10	BE026161	BE026161 db33b09.y
18.2	67.4	802	13	CA164304	CA164304 SCSBR311
18.2	67.4	1101	29	CNS055CF	AL321864 Tetraodon
18	66.7	286	10	B8563010	B8563010 B8563010
18	66.7	457	28	AQ557945	AQ557945 HS_2067.B
18	66.7	490	28	AZ665932	AZ665932 1M0547H12
18	66.7	506	12	BG851071	BG851071 102403081
18	66.7	612	14	CA260332	CA260332 SCRLRT303
18	66.7	651	13	BX879195	BX879195 BX879195
18	66.7	763	13	BX889022	BX889022 BX889022
18	66.7	771	28	BZ704611	BZ704611 PUCRF027D
18	66.7	777	13	BX872370	BX872370 BX872370
18	66.7	777	13	BX876110	BX876110 BX876110
18	66.7	941	28	CC374178	CC374178 PUHSM61TD
18	66.7	960	28	CC351759	CC351759 PUHRE03TD
18	66.7	966	29	CG038721	CG038721 PUQCQ66TB
18	66.7	970	29	CG038725	CG038725 PUQCQ66TD
18	66.7	1068	29	CNS055WL	AL349308 Tetraodon
18	66.7	1130	10	BE729902	BE729902 Tetraodon
17.8	65.9	116	14	CD018332	CD018332 NXLV_019
17.8	65.9	629	14	CB015215	CB015215 Hc_d11_33
17.6	65.2	86	10	B8530332	B8530332 M770045TM
17.6	65.2	189	11	CNS03D52	BX053106 Single re
17.6	65.2	241	11	CNS08QTL	BX024181 Single re
17.6	65.2	264	9	AV263647	AV263647 AV263647
17.6	65.2	338	13	BQ334602	BQ334602 IL2-MT017
17.6	65.2	356	11	CNS0915S	BX059612 Single re
17.6	65.2	359	10	BF912724	BF912724 PM1-UT010
17.6	65.2	391	10	BF935766	BF935766 IL2-NT020
17.6	65.2	396	13	B0790904	B0790904 E4757_Chi
17.6	65.2	437	10	B9330950	B9330950 IL2-NT020
17.6	65.2	445	11	CNS08HLT	BX012377 Single re
17.6	65.2	446	11	CNS08XFO	BX032728 Single re
17.6	65.2	482	11	CNS08MJH	BX018633 Single re
17.6	65.2	510	9	AI179354	AI179354 EST223051
17.6	65.2	567	11	CNS08WT0	BX031936 Single re
17.6	65.2	574	11	CNS09RSO	BX072100 Single re
17.6	65.2	575	11	CNS09AXT	BX050253 Single re
17.6	65.2	695	14	CD837196	CD837196 BN45_051G
17.6	65.2	720	29	CNS01JCN	AI146872 Anopheles
17.6	65.2	824	12	BG618066	BG618066 602644958
17.6	65.2	827	11	CNS08WHX	BX031537 Single re
17.6	65.2	846	29	CC579184	CC579184 CH240_458
17.6	65.2	866	11	CNS03PES	BX068985 Single re
17.6	65.2	911	11	CNS08S91	BX026033 Single re
17.6	65.2	919	11	CNS08VLA	BX030362 Single re
17.6	65.2	940	28	BZ562566	BZ562566 pacs2-164
17.6	65.2	960	14	CB209167	CB209167 AGENCOURT
17.6	65.2	994	10	BF843282	BF843282 PM0-HT107
17.4	64.4	309	29	CE286481	CE286481 tigr-gss
17.4	64.4	353	28	BH016227	BH016227 TDCUJ76TH
17.4	64.4	386	29	CC892346	CC892346 ZMMBBC052
17.4	64.4	404	12	EM092242	EM092242 sah10b03.
17.4	64.4	405	12	EM091628	EM091628 sag99e01.
17.4	64.4	441	10	BB864469	BB864469 BB864469
17.4	64.4	484	13	BU061674	BU061674 Fgr_10_J0
17.4	64.4	517	28	BH401620	BH401620 AG-ND-124
17.4	64.4	527	10	BG044350	BG044350 sac27c10.
17.4	64.4	557	10	BF903516	BF903516 MRI-MT028



[illegible]

prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',  
GAGAGAGAGCGGCGCACTCGAGTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5',  
GAGAGAGATCTCGAGTTCTAAATTAATTCCTCCCTCCCTCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pluescript KS(+) after bulk excision from Lambda FLC I."

## ORIGIN

Query Match 74.8%; Score 20.2; DB 10; Length 324;  
Best Local Similarity 88.0%; Pred. No. 2.1e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCT 25  
|||||  
Db 36 CACCCGCGCTTTAGGATCGACACCT 60  
|||||

## RESULT 2

CA055313/c  
LOCUS 630 bp mRNA linear EST 04-MAR-2003  
DEFINITION ssalrga510.91 mixed\_tissue Salmo salar cDNA, mRNA sequence.

CA055313  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

'Salmo salar (Atlantic salmon)

Salmo salar  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.

REFERENCE 1 (bases 1 to 630)  
GRASP Consortium, Davidson, W.S., Koop, B.F. and  
http://web.uvic.ca/cbr/grasp.

TITLE A survey of Salmo salar transcripts from high complexity cDNA libraries

JOURNAL Unpublished (2002)

COMMENT Contact: Koop BF  
Centre for Biomedical Research  
University of Victoria  
PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada

Tel: 250 472 4067

Fax: 250 472 4075

Email: bkoop@uvic.ca

Genome Sciences Centre, BC Cancer Agency

cDNA preparation, sequencing and bioinformatics:

y Butterfield, R Kirkpatrick, J Asano, N Girn, R Guin, D Lee,  
S Lee, T Olson, P Pandoh, A Prabhu, D Smalhus, L Spence, J Stott,  
S Taylor, G Yang, J Schein, S Jones and M Marra.

## FEATURES

source

1..630  
/organism="Salmo salar"  
/mol\_type="mRNA"  
/strain="McConnell"  
/db\_xref="taxon:8030"  
/clone\_lib="mixed tissue"

/note="Vector: pCMVSPORT6; Library Creator: Research Genetics; Atlantic salmon tissue contributors: Carlo Biagi, Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton, B.C.), Rachel Roper (University of Victoria)"

## ORIGIN

Query Match 72.6%; Score 19.6; DB 13; Length 630;  
Best Local Similarity 84.8%; Pred. No. 4.6e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTGA 27  
|||||  
Db 348 ATCTAGCTTTAGCTCGACACCTGA 323  
|||||

## RESULT 3

CA0259762/c

LOCUS

DEFINITION

nbxb0023D22f CUGI Rice BAC Library Oryza sativa (japonica cultivar-group) genomic clone nbxb0023D22f, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzoideae; Oryza.

REFERENCE 1 (bases 1 to 634)

A BAC End Sequencing Framework to Sequence the Rice Genome

Unpublished (1998)

COMMENT

Contact: Wing RA

Clemson University

Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: TAATACGACTCACTATAGG

Class: BAC ends

High quality sequence stop: 273.

Location/Qualifiers

1..634

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="genomic DNA"

/strain="japonica"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/clone="nbxb0023D22f"

/tissue\_type="Leaf"

/lab\_host="E. coli DH10B"

/clone\_lib="CUGI Rice BAC Library"

/note="Vector: pBelBAC11; Site 1: HindIII; Site 2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

## ORIGIN

Query Match 71.8%; Score 19.4; DB 28; Length 634;

Best Local Similarity 95.2%; Pred. No. 5.6e+02;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 GCGCTTTAGGATCGACACCTG 26  
|||||

Db 294 GCCTTTAGGATCGACACCTG 274  
|||||



```

RESULT 4
BQ683442      1130 bp      mRNA      linear      EST 15-JUL-2002
LOCUS         AGENCOURT_8209478 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6265486
DEFINITION    5', mRNA sequence.
ACCESSION     BQ683442
VERSION       BQ683442.1 GI:21796121
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 1130)
AUTHORS      NIH-MGC http://mgi.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgabbs-remail.nih.gov
              Tissue Procurement: DCTD/DPF
              cDNA Library Preparation: Rubin Laboratory
              DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone Distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCW2432 row: k column: 23
              High quality sequence start: 14
              High quality sequence stop: 246.
              Location/Qualifiers
FEATURES             source
    source
    1..1130
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:6265486"
        /tissue_type="melanotic melanoma, cell line"
        /lab_host="DH10B (phage-resistant)"
        /clone_lib="NIH_MGC_112"
        /note="Organ: skin; Vector: pOTB7; Site: 1: XhoI; Site 2:
        EcoRI; cDNA made by oligo-dT priming. Directionally cloned
        into EcORI/XhoI sites using the following 5' adaptor:
        GGCACGAG(G). Library constructed by Ling Hong in the
        laboratory of Gerald M. Rubin (University of California,
        Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
        Superscript II RT (Life Technologies). Note: this is a
        NIH_MGC Library."
ORIGIN
Query Match      71.1%; Score 19,2; DB 13; Length 1130;
Best Local Similarity 87.5%; Pred. No. 8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACC 24
|||||
Db 1106 CACCGCGCTTTAGGACCCACC 1129

RESULT 5
AZ143037/c     483 bp      DNA      linear      GSS 28-AUG-2000
LOCUS         SP_0041_B1_D04 SP6E Strongylocentrotus purpuratus, purple sea
DEFINITION    urchin, sperm genomic BAC library Strongylocentrotus purpuratus
              genomic clone Plate=41 Col=7 Row=H, genomic survey sequence.
ACCESSION     AZ143037
VERSION       AZ143037.1 GI:8294940
KEYWORDS      GSS.
SOURCE        Strongylocentrotus purpuratus
ORGANISM      Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE     1 (bases 1 to 483)
AUTHORS      Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,

```

```

Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T.,
Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H.
and Hood,L.
A sea urchin genome project: Sequence scan, virtual map, and
additional resources
Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
20402566
MEDLINE
PUBMED
10920195
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 41 row: H column: 7
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 483.
Location/Qualifiers
FEATURES             source
    source
    1..483
        /organism="Strongylocentrotus purpuratus"
        /mol_type="genomic DNA"
        /db_xref="taxon:7668"
        /clone_lib="Strongylocentrotus purpuratus, purple sea
        urchin, sperm genomic BAC library"
        /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
        DH10B"
ORIGIN
Query Match      70.4%; Score 19; DB 28; Length 483;
Best Local Similarity 81.5%; Pred. No. 7.8e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTCA 27
|||||
Db 271 CATCCGCGCTTTAAGCAGCAACCTCA 245

RESULT 6
CE809540      533 bp      DNA      linear      GSS 30-SEP-2003
LOCUS         tigr-gss-dog-1700033511007 Dog Library Canis familiaris genomic,
DEFINITION    genomic survey sequence.
ACCESSION     CE809540
VERSION       CE809540.1 GI:37150524
KEYWORDS      GSS.
SOURCE        Canis familiaris (dog)
ORGANISM      Canis familiaris
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Bukaryota; Metazoa; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE     1 (bases 1 to 533)
AUTHORS      Kirkness,E.F., Batna,V., Halpern,A.L., Levy,S., Remington,K.,
              Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
              Venter,J.C.
              The dog genome: survey sequencing and comparative analysis
              Science 301 (5641), 1898-1903 (2003)
              22875432
              MEDLINE
              14512627
              PUBMED
              14512627
              COMMENT      Contact: Kirkness EF
              The Institute for Genomic Research
              Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
              Rockville, MD 20850, USA
              Tel: 301-838-0200
              Fax: 301-838-0208
              Email: ekirknes@tigr.org
              Class: shotgun.
              Location/Qualifiers
FEATURES             source
    source
    1..533
        /organism="Canis familiaris"
        /mol_type="genomic DNA"
        /strain="Standard Poodle"

```

/db\_xref="taxon:9615"  
/clone\_lib="Dog Library"  
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"

ORIGIN

Query Match 70.4%; Score 19; DB 29; Length 533;  
Best Local Similarity 81.5%; Pred. No. 8e+02; 5; Indels 0; Gaps 0;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27  
|||||  
DB 121 CACCCGCGCTTTGGATAGACACCCGA 147  
|||||

RESULT 7  
BQ973570  
LOCUS  
DEFINITION BQ973570 297 bp mRNA linear EST 21-AUG-2002  
clone QH113D08, mRNA sequence.  
ACCESSION BQ973570  
VERSION BQ973570  
KEYWORDS BQ973570.1 GI:22391093  
SOURCE EST.  
ORGANISM Helianthus annuus (common sunflower)

REFERENCE  
AUTHORS Koziak, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.  
TITLE Lettuce and Sunflower ESTs from the Composite Genome Project  
JOURNAL Unpublished (2002)  
COMMENT Contact: Alexander Kozik [R.W.Michelmore] Department of Vegetable Crops, R.W.Michelmore Lab University of California at Davis (UCD) Asmudson Hall, UCD, Davis, CA 95616, USA Tel: 1-(530)-742-1742 Fax: 1-(530)-752-9659 Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu] belongs to contig QH\_CA\_Contig3348, see http://cgpdb.ucdavis.edu/ for details.  
Plate: QH13 row: D column: 08.

FEATURES  
source  
1..297  
/organism="Helianthus annuus"  
/mol\_type="mRNA"  
/cultivar="RHA801"  
/db\_xref="taxon:4232"  
/clone="QH13D08"  
/lab\_host="E.coli"  
/clone\_lib="QH ABCDI sunflower RHA801"  
/note="Vector: pBRCDNA5flab; The library was constructed from 11 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/TAG\_SEQ=Not found"

ORIGIN  
Query Match 68.9%; Score 18.6; DB 13; Length 297;  
Best Local Similarity 84.0%; Pred. No. 1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCT 25  
|||||

DB 187 CCCCCGAATTTAGGATCGACACCT 211  
|||||

RESULT 8  
BFI49651  
LOCUS  
DEFINITION BFI49651 443 bp mRNA linear EST 29-DEC-2000  
uv67e01.v1 McCarrey Eddy round spermatid Mus musculus cDNA clone  
IWAGS:3664632 5' similar to SW:ACT1\_ACACA P02578 ACTIN 1.; mRNA  
sequence.  
ACCESSION BFI49651  
VERSION BFI49651  
KEYWORDS BFI49651.1 GI:11031046  
SOURCE Mus musculus (house mouse)

REFERENCE  
AUTHORS Marr, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
TITLE The WashU-NCI Mouse EST Project 1999  
JOURNAL Unpublished (1999)  
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mousees@wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MWI:1425400

FEATURES  
source  
1..443  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CD-1"  
/db\_xref="taxon:10090"  
/clone="IWAGS:3664632"  
/sex="male"  
/tissue\_type="round spermatids, pooled from multiple mice"  
/dev\_stage="60 day"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="McCarrey Eddy round spermatid"  
/note="Organ: testis; Vector: pBluescript SK+ (Stratagene); Site 1: XhoII; Site 2: EcoRI; cDNA oligo dt-primed [5'-(GA)10-ACTAGTCGAGTTTCTTTT-3'] and directionally cloned using 5' linkers 5'-AATCGCACGAG-3' and 5'-CTCGTCCG-3'. Size selection of 400bp material gives average insert size ranging from 1-2 kb. Library was mass excised (from lambda-UniZAP-XR) and resulting single-stranded phagemids were prepped and transformed into DH10B. Library contains 98.5% recombinants. References: J. Androl. 20:635-639 and Gene 25:263-269. Library constructed and donated by J. McCarrey, Ph.D. (Southwest Foundation for Biomedical Research, Dept. of Genetics); excision done by E.M. Eddy, Ph.D. (National Institutes of Health, National Institute of Environmental Health Sciences). Original lambda-based library is available through ATCC, catalog #63423."

ORIGIN  
Query Match 68.9%; Score 18.6; DB 10; Length 443;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CCCCCGCTTTAGGATCGACACCTGA 27  
|||||

DB 237 CCCCCGCTTTAGGATCGACACCTGA 261  
|||||

```

http://image.llnl.gov
Plate: LLCM1482 row: h column: 06
High quality sequence shot: 299.
Location/Qualifiers
1. .562
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:469877"
/issue_type="adenocarcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 60"
/note="Organ: pXoslate; Vector: pDNR-LTB (Clontech) ;
Site 1: Sfil (ggcccccctggcc) ; Site 2: Sfil
(ggcattatggcc) ; Double-stranded cDNA was prepared from
cell line RNA. 5' and 3' adaptors were used in cloning as
follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and
3' adaptor sequence:
5'-ATTCTAGAGCGCGAGCGCGACATG-TT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.5
kb (range 0.9-4.0 kb). 14/15 clones contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."

```

## ORIGIN

Query Match	68.9%;	Score 18.6;	DB 12;	Length 662;
Best Local Similarity	84.0%;	Pred. No. 1.3e+03;		
Matches 21; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;
Qy	2	ACCCGGCGCTTTAGGATCGACACCTG	26	
Db	429	ACAACAGCCTTTAGGAGCCACACTG	405	

## RESULT 11

CNS08HL2	LOCUS	719 bp	linear	HTC 07-JAN-2003
CNS08HL2	DEFINITION	Single read from an extremity of a full-length cDNA clone made from Anopheles gambiae total adult females. 3-PRIME end of clone FK0AA19BG04 of strain 6-9 of Anopheles gambiae (African malaria mosquito).		

msqurco?  
ACCESSION BX012210  
VERSION BX012210.1 GI:27561430

KEYWORDS	HTC.
SOURCE	Anopheles gambiae (African malaria mosquito)
ORGANISM	Anopheles gambiae

ORGANISM

## REFERENCES

**AUTHORS**  
Genoscope.

**TITLE**  
Direct Submission

**JOURNAL**  
Submitted (06-JAN-2003) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [segret@genoscope.cns.fr](mailto:segret@genoscope.cns.fr)  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

**FEATURES**  
Location/Qualifiers  
- web : [www.geoscope.cnrs.fr](http://www.geoscope.cnrs.fr)

```

1...719
/organism="Acropheles gambiae"
/mol_type="mRNA"
/strain="6-9"
/db_xref="taxon:7165"
/cclone="FK0AAA1BG04"
/plasmid="pME185-PL"
/note="end : 3-PRIME"

```

## ORIGIN

Query Match	68.9%	Score 18.6;	DB 11;	Length 719;
Best Local Similarity	84.0%;	Pred. No. 1.3e+03;		
Matches 21;	Conservative	0;	Mismatches 4;	Indels 0;
				Gaps 0;

3 CCCGCCCTT<sup>1</sup>AGGATCGACACCTGA 27

RESULT_9	
BJ010761	
LOCUS	BJ010761 503 bp mRNA linear EST 05-DEC-2001
DEFINITION	BJ010761 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA149B11 5', mRNA sequence.
ACCESSION	BJ010761
VERSION	BJ010761.1
KEYWORDS	EST.
SOURCE	Oryzias latipes (Japanese medaka)
ORGANISM	Oryzias latipes
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha; Belontiiformes; Adrainichthyidae; Oryziinae; Oryzias.
REFERENCE	1 (bases 1 to 603)
AUTHORS	Kohara, Y., Shin-i, T., Kimura, T., Navita, T., Jindo, T. and Takeda, H.
TITLE	Medaka EST Project in Takeda's lab
JOURNAL	Unpublished (2001)
COMMENT	Contact: Tadasu Shin-i National Institute of Genetics 1111 Yats, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp.

FEATURES  
source

```

/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-rk"
/db_xref="taxon:8090"
/clone="MF01SSA149B11"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSA cDNA"

```

## ORIGIN

Query Match	68.9%;	Score 18.6;	DB 12;	Length 603;
Best Local Similarity	84.0%;	Pred. No. 1.2e+03;		
Matches 21; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;
QY	3	CCGCGCGTTTAGGATCGACACCTGA	27	
Db	261	CCCGACATTGAGGATCGACACCTGA	285	

RESULT 10					
BG500183/c					
LOCUS	BG500183	662 bp	mRNA	linear	EST 27-MAR-2001
DEFINITION	G0254762Df1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4669877 5'				
					mRNA sequence

ACCESSION	BG500183	
VERSION	BG500183.1	GI:13461700

VERSION	EST.1
KEYWORDS	EST.
SOURCE	Homo sapiens (human)

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	

REFERENCE  
AUTHORS  
1. (bases 1 to 562)  
NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)

COMMENT : Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: CLONETECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

Db 325 CCGCGCTTTACGGTCGACGCTTA 349

CE802854 757 bp DNA linear GSS 30-SEP-2003  
tigr-gss-dog-17000331449669 Dog Library Canis familiaris genomic,  
genomic survey sequence.

CE802854 1 GI:37143776  
GSS.  
Canis familiaris (dog)  
Canis familiaris  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
1 (bases 1 to 757)  
Kirkness, E.F., Batina, V., Halpern, A.L., Levy, S., Remington, K.,  
Rusch, D.B., Delcher, A.L., Pop, M., Wang, M., Fraser, C.M. and  
Venter, J.C.  
The dog genome: survey sequencing and comparative analysis  
Science 301 (5641), 1898-1903 (2003)  
22875432  
14512627  
Contact: Kirkness EF  
The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
Rockville, MD 20850, USA  
Tel: 301-838-0200  
Fax: 301-838-0208  
Email: ekirknes@tigr.org  
Class: shotgun.

FEATURES  
source  
Location/Qualifiers  
1..757  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/strain="Standard Poodle"  
/db\_xref="taxon:9615"  
/clone\_lib="Dog Library"  
/note="Site 1: BstXI; Libraries were prepared from  
peripheral blood"

ORIGIN  
Query Match 68.9%; Score 18.6; DB 29; Length 757;  
Best Local Similarity 84.0%; Pred. No. 1.3e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CCGCGCTTTAGGATCGACACCTGA 27  
|||||  
Db 474 CCGGACCTTAGAATCGACACCTGA 450  
|||||

RESULT 13  
BF789780/c  
LOCUS  
DEFINITION  
BF789780 922 bp mRNA linear EST 12-JAN-2001  
602105331F1 NCI CGAP\_Kid14 Mus musculus cDNA clone IMAGE:4223430  
5', mRNA sequence.

ACCESSION  
BF789780  
VERSION  
BF789780.1 GI:12094816  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 922)  
NIH-MGC <http://mgs.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA library preparation: Life Technologies, Inc.  
cDNA library arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

FEATURES  
source  
Location/Qualifiers  
1..1131  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
<1..2131  
/locus\_tag="HCM1653"

ORIGIN  
Query Match 68.9%; Score 18.6; DB 29; Length 1131;  
Best Local Similarity 84.0%; Pred. No. 1.5e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM9811 row: n column: 07  
High quality sequence stop: 652.  
Location/Qualifiers  
1..922  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4223430"  
/lab\_host="DR10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP Kid14"  
/notes="Organ: Kidney; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.75 kb. Constructed by Life  
Technologies. Note: this is a NCI CGAP Library." |"

ORIGIN  
Query Match 68.9%; Score 18.6; DB 10; Length 922;  
Best Local Similarity 84.0%; Pred. No. 1.4e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ACCGGCTTTAGGATCGACACCTG 26  
|||||  
Db 895 ACCCGCTTTAGGAACGCCCTG 871  
|||||

RESULT 14  
AY403727  
LOCUS  
DEFINITION  
AY403727 Mus musculus HCM1653 gene, VIRTUAL TRANSCRIPT, partial sequence,  
genomic survey sequence.  
ACCESSION  
AY403727  
VERSION  
AY403727.1 GI:39759710  
KEYWORDS  
GSS.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1131)  
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,  
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
Farrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,  
Adams, M.D. and Cargill, M.  
Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene tricos  
Science 302 (5652), 1960-1963 (2003)  
14671302  
2 (bases 1 to 1131)  
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,  
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
Farrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,  
Adams, M.D. and Cargill, M.  
Direct Submission  
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
This sequence was made by sequencing genomic exons and ordering  
them based on alignment  
Location/Qualifiers  
1..1131  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
<1..2131  
/locus\_tag="HCM1653"

ORIGIN  
Query Match 68.9%; Score 18.6; DB 29; Length 1131;  
Best Local Similarity 84.0%; Pred. No. 1.5e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

[illegible]

Eukaryota; Metazoa; Nematoidea; Chromadorea; Rhabditida; Strongylida; Ancylostomatidae; Ancylostomatidae; Bunostominae; Necator.

# REFERENCE

1 (bases 1 to 257)

# AUTHORS

Blaxter, M.L., Parkinson, J., Whittom, C., Daub, J., Guilian, D., Hall, N., Quayle, M. and Bartrell, B.

# TITLE

Edinburgh University/Sanger Centre Nematode EST Project

# JOURNAL

Unpublished (2000)

# COMMENT

Contact: Blaxter ML  
Institute of Cell, Animal and Population Biology  
University of Edinburgh  
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK.

Tel: +44 131 650 5760

Fax: +44 131 670 5450

Email: mark.blaxter@ed.ac.uk

The library was prepared for Dr. David Pritchard University of Nottingham. Sequencing was performed by the Pathogen Sequencing Unit, Sanger Centre, Cambridge, UK (Neil Hall, Mike Quail and Bart Bartrell). The sequence contained a PolyA tail (trimmed)

# PCR PRIMERS

FORWARD: SAC

BACKWARD: T7PL

Plate: 39 row: H column: 07

Seq primer: SAC

High quality sequence stop: 257.

# FEATURES

## source

1..257

/organism="Necator americanus"

/mol\_type="mRNA"

/db\_xref="taxon:51031"

/clone="Na L3\_39H07"

/sex="Mixed"

/dev\_stage="L3"

/clone\_lib="Necator americanus (parasitic nematode) L3"

/note="Vector: PCMV-PCR vector; Site\_1: EcoRI (5' end); Site\_2: XhoI (3' end); Necator americanus is a human hookworm, responsible for debilitating anaemia. The library was constructed by David Pritchard (University of Nottingham, UK) from mRNA from N.americanus L3 stage (Nottingham strain) maintained in hamsters."

# ORIGIN

Query Match 67.4%; Score 18.2; DB 13; Length 257;

Best Local Similarity 87.0%; Pred. No. 1.5e+03;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

# QY

5 CGCGCTTTAGGATCGACACTGA 27

# Db

179 CAGGCTCTAGGATCGCCACTGA 157

# RESULT 17

## BE029790/c

# LOCUS

kp34h02.y1 TBN95TW-SSFH Strongyloides stercoralis cDNA 5' similar

to IR:O61855 O61855 F33E11.4 PROTEIN. ; mRNA sequence.

# ACCESSION

## BE029790

# VERSION

## BE029790.1

# KEYWORDS

## EST.

# SOURCE

## ORGANISM

# REFERENCE

## AUTHORS

McCart, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Willie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarishvili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, N., Waterston, R. and Willson, R.

The Washington Univ. Nematode EST Project, 1999

# TITLE

## JOURNAL

Unpublished (1999)

# COMMENT

Contact: McCarter JP  
The Washington Univ. Nematode EST Project, 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

The library was constructed by Dr. Thomas Nutman and colleagues of NIAID, NIH (tnutman@nih.gov). DNA sequencing by: Washington University Genome Sequencing Center St. Louis.  
Seq primer: -40RP from Gibco  
High quality sequence stop: 450.

# FEATURES

## source

1..510

/organism="Strongyloides stercoralis"

/mol\_type="mRNA"

/strain="Filariform larvae obtained from humans"

/db\_xref="taxon:6248"

/lab\_host="XL-1 Blue MRF" (Stratagene)"

/clone\_lib="TBN95TW-SSFH"

/note="Vector: Lambda Uni-ZAP XR (Stratagene); Site 1: EcoRI; Site 2: XhoI; mRNA was purified from infected filariform larvae which had been isolated from infected humans. cDNA was constructed and, using adaptors, was cloned unidirectionally into the vector from the EcoRI site to the XhoI site. The library has an unamplified titer of 1.5 x 10E6 pfu/ml and an amplified, undiluted titer of 7 x 10E9 pfu/ml. The average insert size of the unamplified library is 975 bp (range, 500-1500)."

# ORIGIN

Query Match 67.4%; Score 18.2; DB 10; Length 510;

Best Local Similarity 87.0%; Pred. No. 1.8e+03;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

# QY

5 CGCGCTTTAGGATCGACACTGA 27

# Db

308 CAGGCTTTAGGATCAACACCAGA 286

# RESULT 18

## BE026161/c

# LOCUS

db33509.y1 Xenopus laevis oocyte Xenopus laevis cDNA clone

XENOPUS SOURCE ID: 5' similar to WP:ZK757.1 CE00467 HYPOTHETICAL

68.7 KD PROTEIN ZK757.1 IN CHROMOSOME III ; mRNA sequence.

# ACCESSION

## BE026161

# VERSION

## BE026161.1

# KEYWORDS

## EST.

# SOURCE

## ORGANISM

# REFERENCE

## AUTHORS

# TITLE

## JOURNAL

# COMMENT

Xenopus laevis (African clawed frog)

Unpublished (1999)

Contact: Sandy Clifton, Ph.D.

WashU Xenopus EST project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library constructed by Bruce Blumberg

Library normalized by Jihwan Song

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: Xenopus clone distribution information for

this library can be found through Research Genetics, visit their

web page at: <http://www.resgen.com/> Please reference the id listed below when ordering this clone: Source lab clone id -  
Seq primer: -40RP from Gibco  
High quality sequence stop: 458.

FEATURES  
source

Location/Qualifiers  
1..542  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="XENOPUS SOURCE ID:"  
/tissue\_type="oocyte (stages 5 and 6)"  
/lab\_host="Top-10 F"  
/clone\_lib="Xenopus laevis oocyte"  
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; cDNA was prepared from 2ug of poly A+ RNA. EcoRI-XhoI cut cDNA was then ligated into UniZap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. SS-library phagemids were prepared by mass excision from the original library and normalized by hybridization to biotinylated driver (prepared from the same library by PCR) to Cot-omega of 11. After removal of hybrids and excess driver by streptavidin sepharose chromatography, the ss-phagemids were made double stranded and electroporated into Top-10 F'. Original library constructed by Bruce Blumberg (Blumberg et al., 1991 Science 253, 194-196; Hawtlet et al., 1995, Genes Dev. 9, 2923-2935). Normalized by Jihwan Song (Song, Cho and Blumberg, unpublished). Note: This is a Xenopus Gene Collection (XGC) library."

ORIGIN

Query Match 67.4%; Score 18.2; DB 10; Length 542;  
Best Local Similarity 87.0%; Pred. No. 1.8e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 CGCGCTTTAGGATCGACACCTGA 27  
|||||  
Db 390 CGCGCTTGAGGATCTCCACCTGA 368  
|||||

RESULT 19  
CA164304

LOCUS  
DEFINITION SCBR23119C04.g RZ3 Saccharum officinarum cDNA clone SCSBR23119C04 5', mRNA sequence.

ACCESSION CA164304  
VERSION CA164304.1 GI:35081583  
KEYWORDS EST.  
SOURCE Saccharum officinarum  
ORGANISM Saccharum officinarum

REFERENCE  
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Saccharum.

TITLE 1 (bases 1 to 802)  
JOURNAL Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.  
COMMENT The libraries that made SUCEST Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

CONTACT: Arruda P  
Centro de Biologia Molecular e Engenharia Genetica  
Universidade Estadual de Campinas  
Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
Tel: 55 19 3788 1137  
Fax: 55 19 3788 1089  
Email: parruda@unicamp.br  
Clone distribution: clone distribution information can be found through the Brazilian Clone Collection Center (BCCC) at <http://www.bcccenter.fcav.unesp.br>  
Plate: 119 row: C column: 04  
Seq primer: T7 Promoter Primer.  
Location/Qualifiers  
1..802  
/organism="Saccharum officinarum"  
/mol\_type="mRNA"

FEATURES  
source

Location/Qualifiers  
1..1101  
/organism="Tetraodon nigroviridis"  
/mol\_type="genomic DNA"

/db\_xref="taxon:4547"  
/clone="SCBR23119C04"  
/lab\_host="DH10B"  
/clone\_lib="RZ3"  
/note="Organ: Shoot-root transition zone from adult plants; Vector: pSport1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from [shoot-root transition zone from adult plants]. cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at <http://sucest.lad.ic.unicamp.br/public>

ORIGIN

Query Match 67.4%; Score 18.2; DB 13; Length 802;  
Best Local Similarity 87.0%; Pred. No. 2e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 CGCGCTTTAGGATCGACACCTG 26  
|||||  
Db 777 CGCGCTTTAGGCGCGACACCTG 799  
|||||

RESULT 20  
CNS055CF

LOCUS  
DEFINITION CNS055CF 1101 bp DNA linear GSS 26-JUL-2000  
Tetraodon nigroviridis genome survey sequence T3 end of clone 030A22 of library A from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL321864  
VERSION AL321864.1 GI:9554748  
KEYWORDS GSS; genome survey sequence.  
SOURCE Tetraodon nigroviridis  
ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.  
1  
Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.  
Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence  
Nat. Genet. 25 (2), 235-238 (2000)  
20296633  
10835645

REFERENCE  
AUTHORS

TITLE

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

TITLE

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
JOURNAL  
TITLE  
JOURNAL

COMMENT

This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES  
source

Location/Qualifiers  
1..1101  
/organism="Tetraodon nigroviridis"  
/mol\_type="genomic DNA"

```

/db_xref="taxon:99883"
/clone="030A22"
/clone_lib="A"
/note="Genoscope sequence ID : COA030B11A1-end : T3"

ORIGIN
Query Match      67.4%; Score 18.2; DB 29; Length 1101;
Best Local Similarity 83.3%; Pred. NO. 2.2e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CCGCGCTTTAGGATCGACACCTG 26
    |||||
Db 848 CCGCTCTTTGGATCNCACCTG 871

RESULT 21
BB563010/c
LOCUS BB563010 RIKEN full-length enriched, adult male heart Mus musculus
DEFINITION CDNA clone 1010001L20 5', mRNA sequence.
ACCESSION BB563010
VERSION BB563010.1 GI:11453902
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 286)
Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,
Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T.,
Hodoyama,Y., Imokani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,
Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y.,
Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K.,
Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,
Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermotabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2): 520-524 (1998)
Itoh,M., Kiteunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5): 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES
Source
1. .286
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="1010001L20"
/sex="male"
/tissue_type="heart"
/dev_stage="adult"

```

```

/lab_host="BNV132"
/clone_lib="RIKEN full-length enriched, adult male heart"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCACTCGAGTGTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGAGATCCAGAGCTCAATTAATTAACCCCCCCCCC 3'].
cDNA was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN
Query Match      66.7%; Score 18; DB 10; Length 286;
Best Local Similarity 80.8%; Pred. No. 1.9e+03;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTGA 27
    |||||
Db 40 AGCGCGCTTTAGGACCTTCACCAGA 15

RESULT 22
AQ557945
LOCUS AQ557945 457 bp DNA linear GSS 29-MAY-1999
DEFINITION HS 2067 B1.F06.T7C CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2067 Col=11 Row=L, genomic survey
sequence.
ACCESSION AQ557945
VERSION AQ557945.1 GI:4917677
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 457)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
93380589
10449764
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com)
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2067 row: L column: 11
Seq primer: T7
Class: BAC ends
High quality sequence stop: 457.
Location/Qualifiers
1. .457
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="plate=2067 Col=11 Row=L"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-Coli DH10B"

```



## ORIGIN

Query Match 66.7%; Score 18; DB 28; Length 457;  
 Best Local Similarity 80.8%; Pred. No. 2.1e+03;  
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTG 26

DB 274 CACCGCGCTTACAGATCCACACCTG 299

## RESULT 23

AZ665932/c

## LOCUS

DEFINITION 1M0547H12R Mouse 10kb plasmid UUC1M library Mus musculus genomic clone UUC1M0547H12 R, genomic survey sequence.

ACCESSION AZ665932

VERSION AZ665932.1

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

## ORGANISM

REFERENCE 1 (bases 1 to 490)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL plasmid inserts

COMMENT Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: dunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0547 row: H column: 12

Seq primer: CACACGAGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 490.

## FEATURES

source

1. 490  
 /location=Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clones="UUC1M0547H12"  
 /sex="Male"

/lab\_hosts="E. Coli strain XL10-Gold, Tl-resistant, P-"  
 /clone\_lib="Mouse 10kb plasmid UUC1M library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 66.7%; Score 18; DB 28; Length 490;  
 Best Local Similarity 80.8%; Pred. No. 2.2e+03;  
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTG 26

DB 208 CACCGCGCTTAGGCTCCACACATG 183

## RESULT 24

BG851071

## LOCUS

DEFINITION 1024030B12.x1 C. reinhardtii CC-1690, normalized, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BG851071

VERSION BG851071.1

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii

ORGANISM Chlamydomonas reinhardtii

REFERENCE 1 (bases 1 to 506)  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.

## AUTHORS

Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P., McDermott, J. P., Silflow, C., Stern, D. and Surzycki, R.

TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants; project phase 2

JOURNAL Unpublished (2000)

COMMENT Contact: Charles Hauser

DCMB Box 91000

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

Location/Qualifiers

1. 506  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="CC-1690 wild type mt+ 21gr"  
 /db\_xref="taxon:3055"  
 /clone\_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"

/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. polyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

## ORIGIN

Query Match 66.7%; Score 18; DB 12; Length 506;  
 Best Local Similarity 80.8%; Pred. No. 2.2e+03;  
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTGA 27

DB 205 ACACGAGCTTTGGACGACCTGACCTGA 230

## RESULT 25

CA260332

## LOCUS

DEFINITION SCLRT3034E08.g RT3 Saccharum officinarum cDNA clone SCLRT3034E08



```

/clone="tcbk0045c.a.11"
/tissue type="multi-tissues"
/dev stage="from embryos to adults"
/lab host="DH10B"
/clone lib="AGENAE Rainbow trout multi-tissues-normalized
(tcbbk)"
/notes="vector: pT73D-pac; Rainbow trout
multi-tissues-normalized + 2 subtractions; Clone
distribution : AGENAE Resource centre. Francois PIUMI,
Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
Etude du genome (LREG), Domaine de Vilvert, 78352,
Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
(0) 1.34.65.22.73"

ORIGIN
Query Match 66.7%; Score 18; DB 13; Length 763;
Best Local Similarity 80.8%; Pred. No. 2.4e+03;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACCCGCGCTTTAGGATCGACACCTGA 27
Db 435 ATCTAGCTTTAGCATGCCACCTGA 460

```

```

RESULT 28
BZ704611
LOCUS BZ704611 771 bp DNA linear GSS 19-FEB-2003
DEFINITION PUCF02TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTal26A04,
genomic survey sequence.
ACCESSION BZ704611
VERSION BZ704611.1 GI:28425019
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

```

```

REFERENCE 1 (bases 1 to 771)
AUTHORS Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
Maize Genomics Consortium
Unpublished (2003)
CONTACT: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: Tg
Class: sheared ends.
Location/Qualifiers
1 . 771
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTal26A04"
/clone lib="ZM 0.6_1.0_KB"
/notes="vector: pCR4-ToPO; site 1: EcoRI; 0.6-1.0 kb high
cot selected genomic DNA library"

```

```

ORIGIN
Query Match 66.7%; Score 18; DB 28; Length 771;
Best Local Similarity 80.8%; Pred. No. 2.4e+03;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CACCGCGCTTTAGGATCGACACCTG 26
Db 230 CACCGCGGATTAGATAGAAACCTG 255

```

```

RESULT 29

```

```

BX872970
LOCUS BX872970 777 bp mRNA linear EST 17-DEC-2003
DEFINITION Oncorhynchus mykiss cDNA clone tcbk0025c.d.13 5prim, mRNA sequence.
ACCESSION BX872970
VERSION BX872970.1 GI:40001515
KEYWORDS EST.
SOURCE Oncorhynchus mykiss (rainbow trout)
ORGANISM Oncorhynchus mykiss
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
1 (bases 1 to 777)

```

```

REFERENCE 1 (bases 1 to 777)
AUTHORS Govoroun,M., Guiguen,Y. and Le Gac,F.
TITLE Construction and primary characterization of normalized cDNA
libraries in rainbow trout, Oncorhynchus mykiss
JOURNAL Unpublished (2003)
COMMENT Contact: Guiguen Y
INRA - SCRIBE
Campus de Beaulieu, RENNES cedex, 35042, France
Tel: 02.23.48.50.09
Fax: 02.23.48.50.20
Email: Yann.Guiguen@beaulieu.rennes.inra.fr
Sequence cleaned of vector, adaptor and repetitions. Contact us
at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
sequence.
Plate: 0025 row: d column: 13
Seq primer: M13R.
Location/Qualifiers
1 . 777
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="tcbk0025c.d.13"
/tissue type="multi-tissues"
/dev stage="from embryos to adults"
/lab host="DH10B"
/clone lib="AGENAE Rainbow trout multi-tissues-normalized
(tcbbk)"
/notes="vector: pT73D-pac; Rainbow trout
multi-tissues-normalized + 2 subtractions; Clone
distribution : AGENAE Resource centre. Francois PIUMI,
Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
Etude du genome (LREG), Domaine de Vilvert, 78352,
Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
(0) 1.34.65.22.73"

```

```

FEATURES
source

```

```

ORIGIN

```

```

Query Match 66.7%; Score 18; DB 13; Length 777;
Best Local Similarity 80.8%; Pred. No. 2.4e+03;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACCCGCGCTTTAGGATCGACACCTGA 27
Db 582 ATCTAGCTTTAGCATGCCACCTGA 607

```

```

RESULT 30
BX876110
LOCUS BX876110 777 bp mRNA linear EST 17-DEC-2003
DEFINITION Oncorhynchus mykiss cDNA clone tcbk0014c.f.23 5prim, mRNA sequence.
ACCESSION BX876110
VERSION BX876110.1 GI:40004655
KEYWORDS EST.
SOURCE Oncorhynchus mykiss (rainbow trout)
ORGANISM Oncorhynchus mykiss
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
1 (bases 1 to 777)

```

```

REFERENCE 1 (bases 1 to 777)
AUTHORS Govoroun,M., Guiguen,Y. and Le Gac,F.
TITLE Construction and primary characterization of normalized cDNA
libraries in rainbow trout, Oncorhynchus mykiss
JOURNAL Unpublished (2003)
COMMENT Contact: Guiguen Y
INRA - SCRIBE
Campus de Beaulieu, RENNES cedex, 35042, France
Tel: 02.23.48.50.09
Fax: 02.23.48.50.20
Email: Yann.Guiguen@beaulieu.rennes.inra.fr
Sequence cleaned of vector, adaptor and repetitions. Contact us
at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
sequence.
Plate: 0025 row: d column: 13
Seq primer: M13R.
Location/Qualifiers
1 . 777
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="tcbk0025c.d.13"
/tissue type="multi-tissues"
/dev stage="from embryos to adults"
/lab host="DH10B"
/clone lib="AGENAE Rainbow trout multi-tissues-normalized
(tcbbk)"
/notes="vector: pT73D-pac; Rainbow trout
multi-tissues-normalized + 2 subtractions; Clone
distribution : AGENAE Resource centre. Francois PIUMI,
Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
Etude du genome (LREG), Domaine de Vilvert, 78352,
Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
(0) 1.34.65.22.73"

```

libraries in rainbow trout, Oncorhynchus mykiss  
Unpublished (2003)  
Contact: Guiguen Y  
INRA - SCRIBE  
Campus de beaulieu, RENNES cedex, 35042, France  
Tel: 02.23.48.50.09  
Fax: 02.23.48.50.20  
Email: Yann.Guiguen@beaulieu.rennes.inra.fr  
Sequence cleaned of vector, adaptor and repetitions. Contact us  
at signasupport@jouy.inra.fr to obtain the chromatogram of this  
sequence.  
Plate: 0014 row: f column: 23  
Seq primer: M13R.

FEATURES  
source  
1..777  
Location/Qualifiers  
/organism="Oncorhynchus mykiss"  
/mol\_type="mRNA"  
/db\_xref="taxon:8022"  
/clone="tcbk0014c.f.23"  
/tissue\_type="multi-tissues"  
/dev\_stage="from embryos to adults"  
/lab\_host="DH10B"  
/clone\_lib="AGENAE Rainbow trout multi-tissues-normalized  
(tcbk)"  
/note="Vector: pT73D-pac; Rainbow trout  
multi-tissues-normalized + 2 subtractions; Clone  
distribution : AGENAE Resource centre. Francois PIUMI,  
Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et  
Etude du genome (LREG), Domaine de Vilvert, 78352,  
Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33  
(0) 1.34.65.22.73"

ORIGIN

Query Match 66.7%; Score 18; DB 13; Length 777;  
Best Local Similarity 80.8%; Pred. No. 2.4e+03;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACCCGCGCTTTAGGATCGACACCTGA 27  
Db 582 ATCTAGCTTTAGCATGCCACCTGA 607

Search completed: June 20, 2004, 14:13:35  
Job time : 1288.71 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 152.714 Seconds  
(without alignments)  
809.940 Million cell updates/sec

Title: US-10-624-714-12

Perfect score: 27

Sequence: 1 caccgcgcttagatcgacacctga 27

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034952

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Published Applications NA:

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:
- 12: /cgn2\_6/ptodata/2/pubpna/US09D\_NEW\_PUB.seq:
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:
- 17: /cgn2\_6/ptodata/2/pubpna/US10D\_NEW\_PUB.seq:
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.2	71.1	2940917	13	US-10-027-632-174763
2	19.2	71.1	2940917	16	US-10-027-632-174763
3	18.6	68.9	2000	9	US-09-938-842A-4597
4	18.6	68.9	2000	11	US-09-938-842A-4597
5	18.6	68.9	2000	9	US-09-938-842A-4597
6	18.6	68.9	2000	11	US-09-938-842A-4597
7	18.6	68.9	2000	9	US-09-938-842A-4597
8	18.6	68.9	2000	11	US-09-938-842A-4597
9	17.6	65.2	894	15	US-10-138-970A-1
10	17.6	65.2	894	15	US-10-138-970A-1
11	17.6	65.2	894	15	US-10-138-970A-1
12	17.6	65.2	894	15	US-10-138-970A-1
13	17.4	64.4	842	13	US-10-027-632-324925
14	17.4	64.4	842	13	US-10-027-632-324925

Sequence 324925,	16	US-10-027-632-324925	842	64.4	17.4	C 15
Sequence 325019,	16	US-10-027-632-325019	842	64.4	17.4	C 16
Sequence 27104, A	16	US-10-369-493-27104	1050	64.4	17.4	C 17
Sequence 103399,	13	US-10-424-599-103399	1063	64.4	17.4	C 18
Sequence 6630, Ap	15	US-10-156-761-6630	1296	64.4	17.4	C 19
Sequence 26417, A	13	US-10-282-122A-26417	1789	64.4	17.4	C 20
Sequence 135, App	9	US-09-712-363-135	2283	64.4	17.4	C 21
Sequence 28726, A	13	US-10-282-122A-28726	2283	64.4	17.4	C 22
Sequence 1937, Ap	9	US-09-974-300-1937	1999	64.4	17.4	C 23
Sequence 23519, A	13	US-10-282-122A-23519	1257	63.7	17.2	C 24
Sequence 45829, A	16	US-10-369-493-45829	2493	63.7	17.2	C 25
Sequence 30891, A	9	US-09-864-761-30891	279	63.0	17	C 26
Sequence 131520,	13	US-10-424-599-131520	323	63.0	17	C 27
Sequence 661, App	9	US-09-728-445-661	349	63.0	17	C 28
Sequence 108794,	13	US-10-424-599-108794	400	63.0	17	C 29
Sequence 14328, A	9	US-09-864-761-14328	463	63.0	17	C 30
Sequence 40188, A	16	US-10-369-493-40188	559	63.0	17	C 31
Sequence 1494, A	13	US-10-282-122A-1494	597	63.0	17	C 32
Sequence 8860, Ap	13	US-10-282-122A-8860	1485	63.0	17	C 33
Sequence 24148, A	13	US-10-282-122A-24148	1674	63.0	17	C 34
Sequence 20169, A	13	US-10-282-122A-20169	1675	63.0	17	C 35
Sequence 43341, A	16	US-10-369-493-43341	2772	63.0	17	C 36
Sequence 156, App	10	US-09-989-442-156	12409	63.0	17	C 37
Sequence 8437, Ap	13	US-09-764-891-8437	8163	62.2	16.8	C 38
Sequence 8101, Ap	265	US-10-085-783A-8101	265	61.5	16.6	C 39
Sequence 8101, Ap	265	US-10-242-535A-8101	265	61.5	16.6	C 40
Sequence 9001, Ap	400	US-10-085-783A-9001	400	61.5	16.6	C 41
Sequence 9001, Ap	400	US-10-242-535A-9001	400	61.5	16.6	C 42
Sequence 259, App	9	US-09-770-444-259	467	61.5	16.6	C 43
Sequence 460, App	9	US-09-887-576-460	546	61.5	16.6	C 44
Sequence 687, App	9	US-09-938-842A-687	546	61.5	16.6	C 45
Sequence 28094, A	11	US-09-938-842A-687	546	61.5	16.6	C 46
Sequence 26133, A	13	US-10-282-122A-28094	939	61.5	16.6	C 47
Sequence 35426, A	16	US-10-369-493-26133	1074	61.5	16.6	C 48
Sequence 41735, A	13	US-10-425-114-35426	1450	61.5	16.6	C 49
Sequence 63, Appl	16	US-10-369-493-41735	1578	61.5	16.6	C 50
Sequence 37320, A	14	US-10-078-929-63	1587	61.5	16.6	C 51
Sequence 11722, A	13	US-10-282-122A-37320	1675	61.5	16.6	C 52
Sequence 39316, A	13	US-10-282-122A-39316	1677	61.5	16.6	C 53
Sequence 115, App	15	US-10-270-333-115	1995	61.5	16.6	C 54
Sequence 78621, A	13	US-10-424-599-78621	5937	60.7	16.4	C 55
Sequence 268282,	13	US-10-027-632-268282	450	60.7	16.4	C 56
Sequence 19476, A	16	US-10-027-632-268282	450	60.7	16.4	C 57
Sequence 1537, Ap	15	US-10-029-386-19476	451	60.7	16.4	C 58
Sequence 1516, Ap	16	US-10-412-699B-1537	543	60.7	16.4	C 59
Sequence 5883, Ap	16	US-10-374-780A-1516	543	60.7	16.4	C 60
Sequence 9923, Ap	15	US-10-029-386-5883	581	60.7	16.4	C 61
Sequence 1574, Ap	9	US-09-864-761-9923	595	60.7	16.4	C 62
Sequence 4757, A	9	US-09-974-300-1574	735	60.7	16.4	C 63
Sequence 39509, A	16	US-10-369-493-4757	894	60.7	16.4	C 64
Sequence 36877, A	13	US-10-282-122A-39509	975	60.7	16.4	C 65
Sequence 19724, A	13	US-10-282-122A-36877	1004	60.7	16.4	C 66
Sequence 26318, A	13	US-10-282-122A-19724	1011	60.7	16.4	C 67
Sequence 15461, A	16	US-10-369-493-26318	1056	60.7	16.4	C 68
Sequence 32483, A	13	US-10-425-114-15461	1078	60.7	16.4	C 69
Sequence 1601, Ap	13	US-10-282-122A-32483	1188	60.7	16.4	C 70
Sequence 1520, Ap	16	US-10-412-699B-1601	1494	60.7	16.4	C 71
Sequence 6112, Ap	13	US-10-374-780A-1520	1494	60.7	16.4	C 72
Sequence 12781, A	13	US-10-156-761-6112	1593	60.7	16.4	C 73
Sequence 1342, Ap	13	US-10-282-122A-12781	2331	60.7	16.4	C 74
Sequence 38, Appl	13	US-10-342-887-1342	2565	60.7	16.4	C 75
Sequence 3707, Ap	10	US-09-927-827-38	2576	60.7	16.4	C 76
Sequence 72, Appl	9	US-09-880-107-3707	2665	60.7	16.4	C 77
Sequence 393, App	15	US-10-252-157-72	2711	60.7	16.4	C 78
Sequence 3, Appl	17	US-10-283-975A-393	2761	60.7	16.4	C 79
Sequence 5, Appl	16	US-09-945-537-3	2912	60.7	16.4	C 80
Sequence 1, Appl	16	US-10-349-670-5	4064	60.7	16.4	C 81
Sequence 182, App	15	US-10-085-959-182	6861	60.7	16.4	C 82
Sequence 1867, Ap	13	US-10-087-192-1867	71132	60.7	16.4	C 83
Sequence 97, Appl	12	US-09-997-722-97	96599	60.7	16.4	C 84





```
; LENGTH: 2940917
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(2940917)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174763

Query Match
Best Local Similarity 71.1%; Score 19.2; DB 16; Length 2940917;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACC 24
   ||||| ||||| ||||| |||||
Db 1793893 CACCCGCGCTTTAGGATCGACACC 1793916

RESULT 3
US-09-938-842A-4597
; Sequence 4597, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4597
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4597

Query Match
Best Local Similarity 68.9%; Score 18.6; DB 9; Length 2000;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACTG 26
   ||||| ||||| ||||| |||||
Db 783 ACCCGCGCTTTAGGATCGACACTG 807

RESULT 4
US-09-938-842A-4597
; Sequence 4597, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4597
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4597

Query Match
Best Local Similarity 84.0%; Score 18.6; DB 9; Length 2000;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACTG 26
   ||||| ||||| ||||| |||||
Db 783 ACCCGCGCTTTAGGATCGACACTG 807

PRIORITY FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4597
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4597

Query Match
Best Local Similarity 84.0%; Score 18.6; DB 11; Length 2000;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACTG 26
   ||||| ||||| ||||| |||||
Db 783 ACCCGCGCTTTAGGATCGACACTG 807

PRIORITY FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4597
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4597

Query Match
Best Local Similarity 84.0%; Score 18.6; DB 9; Length 1287;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACTG 26
   ||||| ||||| ||||| |||||
Db 90 CAGCGCGCTTTAGGATCGACACTG 65

RESULT 6
US-10-627-476-621/c
; Sequence 621, Application US/10627476
; Publication No. US20040030116A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Mark
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schoder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
; TITLE OF INVENTION: TRANSPORT
; FILE REFERENCE: BGI-125CPCN
US-10-627-476-621/c
```



```

; CURRENT APPLICATION NUMBER: US/10/627,476
; CURRENT FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: 09/602,787
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: USN 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931454.3
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931478.0
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931563.9
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932122.1
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932124.8
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932180.9
; PRIOR FILING DATE: 1999-07-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 678
; SEQ ID NO 621
; LENGTH: 1335
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1312)
; OTHER INFORMATION: FXA02233
US-10-627-476-621

Query Match      66.7%; Score 18; DB 13; Length 1335;
Best Local Similarity 80.8%; Pred. No. 84;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTG 26
   |||||
Db 190 CAGCGCTCTTTAGGAGCGCAACTG 165

RESULT 7
US-10-627-476-619/c
; Sequence 619, Application US/10627476
; Publication No. US20040030116A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Mark
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schoder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
; FILE REFERENCE: BGI-125CPCN
; CURRENT APPLICATION NUMBER: US/10/627,476
; CURRENT FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: 09/602,787
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: USN 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931454.3
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931478.0
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931563.9
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932122.1
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932124.8
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932180.9
; PRIOR FILING DATE: 1999-07-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 678
; SEQ ID NO 621
; LENGTH: 1335
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1312)
; OTHER INFORMATION: FXA02233
US-10-627-476-621

Query Match      66.7%; Score 18; DB 13; Length 1335;
Best Local Similarity 80.8%; Pred. No. 84;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTG 26
   |||||
Db 190 CAGCGCTCTTTAGGAGCGCAACTG 165

RESULT 8
US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAYAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match      66.7%; Score 18; DB 9; Length 3309400;
Best Local Similarity 80.8%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTG 26
   |||||
Db 2048517 CAGCGCTCTTTAGGAGCGCAACTG 2048542

RESULT 9
US-10-138-970A-1
; Sequence 1, Application US/10138970A
; Publication No. US20030079255A1
```

```

; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932180.9
; PRIOR FILING DATE: 1999-07-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 678
; SEQ ID NO 619
; LENGTH: 1410
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1387)
; OTHER INFORMATION: RXN02233
US-10-627-476-619

Query Match      66.7%; Score 18; DB 13; Length 1410;
Best Local Similarity 80.8%; Pred. No. 84;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTG 26
   |||||
Db 190 CAGCGCTCTTTAGGAGCGCAACTG 165

RESULT 8
US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAYAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match      66.7%; Score 18; DB 9; Length 3309400;
Best Local Similarity 80.8%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTG 26
   |||||
Db 2048517 CAGCGCTCTTTAGGAGCGCAACTG 2048542

RESULT 9
US-10-138-970A-1
; Sequence 1, Application US/10138970A
; Publication No. US20030079255A1
```



```
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      65.2%; Score 17.6; DB 15; Length 9025608;
Best Local Similarity 83.3%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 CCGCGCGCTTTAGGATCGACACCTG 26
    ||||| ||||| ||||| ||||| |||||
Db 8803965 CCGCGCTTCAGATCGTCACCG 8803988

RESULT 13
US-10-027-632-324925/c
; Sequence 324925, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 324925
; LENGTH: 842
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-324925

Query Match      64.4%; Score 17.4; DB 13; Length 842;
Best Local Similarity 77.8%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCTGA 27
    ||||| ||||| ||||| ||||| |||||
Db 273 CACCCGAGCTTCAGGAACAACAGCTGA 247

RESULT 14
US-10-027-632-325019/c
; Sequence 325019, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 324925
; LENGTH: 842
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-324925

Query Match      64.4%; Score 17.4; DB 13; Length 842;
Best Local Similarity 77.8%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCTGA 27
    ||||| ||||| ||||| ||||| |||||
Db 273 CACCCGAGCTTCAGGAACAACAGCTGA 247

RESULT 15
US-10-027-632-324925/c
; Sequence 324925, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 324925
; LENGTH: 842
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-324925

Query Match      64.4%; Score 17.4; DB 16; Length 842;
Best Local Similarity 77.8%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCTGA 27
    ||||| ||||| ||||| ||||| |||||
Db 273 CACCCGAGCTTCAGGAACAACAGCTGA 247

RESULT 16
US-10-027-632-325019/c
```

; Sequence 325019, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 325019  
; LENGTH: 842  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-325019

Query Match 64.4%; Score 17.4; DB 16; Length 842;  
Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27  
DB 273 CACCCGCGCTTTAGGATCGACACCTGA 247

RESULT 17  
US-10-369-493-27104  
; Sequence 27104, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 27104  
; LENGTH: 1050  
; TYPE: DNA  
; ORGANISM: Neurospora crassa  
US-10-369-493-27104

Query Match 64.4%; Score 17.4; DB 16; Length 1050;  
Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27  
DB 669 CATCCGCGCTTCAGGATGGAGCCAGA 695

RESULT 18  
US-10-424-599-103399/c  
; Sequence 103399, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovacic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 103399  
; LENGTH: 1063  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_64388C.1  
US-10-424-599-103399

Query Match 64.4%; Score 17.4; DB 13; Length 1063;  
Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27  
DB 447 CACCCGCGCTTTAGGATCGACACCTGA 421

RESULT 19  
US-10-156-761-6630/c  
; Sequence 6630, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 6630  
; LENGTH: 1296  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1296)  
US-10-156-761-6630

Query Match 64.4%; Score 17.4; DB 15; Length 1296;  
Best Local Similarity 77.8%; Pred. No. 1.7e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27  
DB 860 CAGCCGCGCTTTAGGATCGACACCTGA 834

RESULT 20  
US-10-282-122A-26417/c

Sequence 26417, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 26417  
LENGTH: 1789  
TYPE: DNA  
ORGANISM: Mycobacterium bovis  
US-10-282-122A-26417

Query Match 64.4%; Score 17.4; DB 13; Length 1789;  
Best Local Similarity 77.8%; Pred. No. 1.7e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27  
DB 1293 CGCCGCGCGCCAGGATCGTCACCTTA 1267

RESULT 21  
US-09-712-363-135/C  
Sequence 135, Application US/09712363  
Patent No. US20020164589A1  
GENERAL INFORMATION:  
APPLICANT: Eisenberg, David  
APPLICANT: Rotstein, Sergio H.  
APPLICANT: Marcotte, Edward M.  
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
FILE REFERENCE: 07419-032001  
CURRENT APPLICATION NUMBER: US/09/712,363  
CURRENT FILING DATE: 2000-11-13  
PRIOR APPLICATION NUMBER: PCT/US00/02246  
PRIOR FILING DATE: 2000-01-28  
PRIOR APPLICATION NUMBER: 60/179,531

PRIOR FILING DATE: 2000-02-01  
PRIOR APPLICATION NUMBER: 60/117,844  
PRIOR FILING DATE: 1999-01-29  
PRIOR APPLICATION NUMBER: 60/118,206,  
PRIOR FILING DATE: 1999-02-01  
PRIOR APPLICATION NUMBER: 60/126,593  
PRIOR FILING DATE: 1999-03-26  
PRIOR APPLICATION NUMBER: 60/134,093  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 60/134,092  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 60/165,124  
PRIOR FILING DATE: 1999-11-12  
PRIOR APPLICATION NUMBER: 60/165,086  
PRIOR FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 292  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 135  
LENGTH: 2283  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-135

Query Match 64.4%; Score 17.4; DB 9; Length 2283;  
Best Local Similarity 77.8%; Pred. No. 1.7e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27  
DB 1355 CGCCGCGCGCCAGGATCGTCACCTTA 1329

RESULT 22  
US-10-282-122A-28726/C  
Sequence 28726, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16

```
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28726
; LENGTH: 2283
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-28726

Query Match      64.4%; Score 17.4; DB 13; Length 2283;
Best Local Similarity 77.8%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      1 CACCCGCGCTTTAGGATCGACACCTGA 27
      |||||
Db      1355 CGCCCGCGCGCCAGGATCGTCACCTTA 1329

RESULT 23
US-09-974-300-1937/c
; Sequence 1937, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1937
; LENGTH: 999
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1937

Query Match      63.7%; Score 17.2; DB 9; Length 999;
Best Local Similarity 86.4%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 CACCCGCGCTTTAGGATCGACA 22
      |||||
Db      960 CACACGCGCTTTAGGATCCATA 939

RESULT 24
US-10-282-122A-23919
; Sequence 23919, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28726
; LENGTH: 2283
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-28726

Query Match      64.4%; Score 17.4; DB 13; Length 2283;
Best Local Similarity 77.8%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      1 CACCCGCGCTTTAGGATCGACACCTGA 27
      |||||
Db      1355 CGCCCGCGCGCCAGGATCGTCACCTTA 1329

RESULT 23
US-09-974-300-1937/c
; Sequence 1937, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1937
; LENGTH: 999
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1937

Query Match      63.7%; Score 17.2; DB 9; Length 999;
Best Local Similarity 86.4%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 CACCCGCGCTTTAGGATCGACA 22
      |||||
Db      960 CACACGCGCTTTAGGATCCATA 939

RESULT 24
US-10-282-122A-23919
; Sequence 23919, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23919
; LENGTH: 1257
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-23919

Query Match      63.7%; Score 17.2; DB 13; Length 1257;
Best Local Similarity 86.4%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 CCGCGCTTTAGGATCGACACCT 25
      |||||
Db      340 CCGCGCTTTGATCGACACCT 361

RESULT 25
US-10-369-493-45829
; Sequence 45829, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 45829
; LENGTH: 2493
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-45829

Query Match      63.7%; Score 17.2; DB 16; Length 2493;
Best Local Similarity 86.4%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      6 GCGCTTTAGGATCGACACCTGA 27
      |||||
Db      1688 GCGCTTTAGGATCGCGCGCTGA 1709

RESULT 26
US-09-864-761-30891/c
; Sequence 30891, Application US/09864761
```

```
Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aescm.ca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 30891
; LENGTH: 279
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005669.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.53
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.44
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.5
; OTHER INFORMATION: SWISSPROT HIT: P23071, EVALUAE 2.40e-01
; OTHER INFORMATION: EST HUMAN HIT: AWS10940.1, EVALUAE 6.40e-02
; OTHER INFORMATION: NT HIT: AJ289880.1, EVALUAE 0.00e+00
US-09-864-761-30891
Query Match 63.0%; Score 17; DB 9; Length 279;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTG 26
Db 36 ACCCTCTCTTTGGGATCGACCACTG 12

RESULT 27
US-10-624-599-131520/c
; Sequence 131520, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 295864
; SEQ ID NO 131520
; LENGTH: 323
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_89770C.1
US-10-424-599-131520
Query Match 63.0%; Score 17; DB 13; Length 323;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTG 26
Db 131 ACCCGCGCTTTAGGATCGACACCTG 107

RESULT 28
US-09-728-445-661
; Sequence 661, Application US/09728445
; Patent No. US20020102543A1
; GENERAL INFORMATION:
; APPLICANT: Friedzich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and
; FILE REFERENCE: LEX-0102-USA
; CURRENT APPLICATION NUMBER: US/09/728,445
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,358
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 891
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 661
; LENGTH: 349
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(349)
; OTHER INFORMATION: n = A,T,C or G
US-09-728-445-661
Query Match 63.0%; Score 17; DB 9; Length 349;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCT 25
Db 158 CACCCGCGCTCTCAGATCGACATCT 182

RESULT 29
US-10-424-599-108794
; Sequence 108794, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
```

APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 108794  
LENGTH: 400  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_69254C.1  
US-10-424-599-108794

Query Match 63.0%; Score 17; DB 13; Length 400;  
Best Local Similarity 80.0%; Pred. No. 2.3e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCT 25  
||| ||||| ||||| ||||| |||||  
Db 19 CAACTGTGTTTGGATCGAAACCT 43

## RESULT 30

US-09-864-761-14328/c  
Sequence 14328, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecm1ca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/532,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
SEQ ID NO 14328  
LENGTH: 463  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC005669.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.53  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.44  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.5  
US-09-864-761-14328

Query Match 63.0%; Score 17; DB 9; Length 463;  
Best Local Similarity 80.0%; Pred. No. 2.4e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACCCGGCGCTTTAGGATCGACACCTG 26  
||| ||||| ||||| ||||| |||||  
Db 262 ACCCTCTCTTTGGATCCCAACTG 238

Search completed: June 20, 2004, 17:45:28  
Job time : 234.714 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 543.069 Seconds  
(without alignments)  
1596.226 Million cell updates/sec

Title: US-10-624-714-13  
Perfect score: 20  
Sequence: 1 ggcgcgcacacactgcactt 20

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 300 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.scs.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.scs.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.atgo.hum.\*
- 40: em.atgo.mus.\*
- 41: em.atgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	20	100.0	335	1	MBO298910	AJ298910 Mycobacte
2	20	100.0	335	1	MBO298911	AJ298911 Mycobacte
3	20	100.0	390	1	MMI298907	AJ298907 Mycobacte
4	20	100.0	390	1	MMI298909	AJ298909 Mycobacte
5	20	100.0	414	1	MBO298905	AJ298905 Mycobacte
6	20	100.0	469	1	MTU298906	AJ298906 Mycobacte
7	20	100.0	493	1	MBO298908	AJ298908 Mycobacte
8	20	100.0	10894	1	AE007121	AE007121 Mycobacte
9	20	100.0	318050	1	EX248344	EX248344 Mycobacte
10	20	100.0	348676	15	EX842581	EX842581 Mycobacte
C 11	18.4	92.0	38477	1	MSGB1723CS	L78825 Mycobacteri
C 12	18.4	92.0	42926	1	MLCB1243	AL023635 Mycobacte
C 13	18.4	92.0	348950	1	MLEPRN7	AL583923 Mycobacte
C 14	17	85.0	19632	1	SNO224512	AJ224512 Streptomy
C 15	17	85.0	67893	8	AP003754	AP003754 Oryza sat
C 16	17	85.0	137312	2	AC135598	AC135598 Oryza sat
C 17	17	85.0	158261	8	CNS0809S	AL772414 Oryza sat
C 18	17	85.0	308015	1	AE016783	AE016783 Pseudomon
C 19	16.8	84.0	276	8	AF087589	AF087589 Saxifraga
C 20	16.8	84.0	420	1	CXSTPDUCL	X87993 C.xyli DNA
C 21	16.8	84.0	858	6	AXI22356	AXI22356 Sequence
C 22	16.8	84.0	858	6	BD164473	BD164473 Novel pol
C 23	16.8	84.0	968	8	AK108178	AK108178 Oryza sat
C 24	16.8	84.0	1654	3	AF038597	AF038597 Drosophil
C 25	16.8	84.0	1733	3	AY069220	AY069220 Drosophil
C 26	16.8	84.0	1755	3	AY071071	AY071071 Drosophil
C 27	16.8	84.0	2163	3	DMU29529	U29529 Drosophila
C 28	16.8	84.0	11366	1	AE011952	AE011952 Xanthomon
C 29	16.8	84.0	64195	2	AP004394	AP004394 Oryza sat
C 30	16.8	84.0	98286	2	AP004394	Continuation (33 o
C 31	16.8	84.0	110000	2	LMFLCHR36	AL049694 Human DNA
C 32	16.8	84.0	123296	9	HSB799B18	AC004642 Drosophil
C 33	16.8	84.0	148432	3	AC004642	AC004642 Drosophil
C 34	16.8	84.0	154840	3	AC099018	AC099018 Drosophil
C 35	16.8	84.0	157851	2	AC020509	AC020509 Drosophil
C 36	16.8	84.0	159478	2	AL359638	AL359638 Homo sapi
C 37	16.8	84.0	173613	3	AC007475	AC007475 Drosophil
C 38	16.8	84.0	184621	3	AC007440	AC007440 Drosophil
C 39	16.8	84.0	204697	2	AC120995	AC120995 Rattus no
C 40	16.8	84.0	222981	10	AL669981	AL669981 Mouse DNA
C 41	16.8	84.0	258635	2	AC106208	AC106208 Rattus no
C 42	16.8	84.0	292919	3	AE003823	AE003823 Drosophil
C 43	16.8	84.0	303823	3	AE003462	AE003462 Drosophil
C 44	16.8	84.0	337200	1	AP005280	AP005280 Corynebac
C 45	16.8	84.0	343243	1	EX640414	EX640414 Bordetell
C 46	16.8	84.0	347894	1	EX640431	EX640431 Bordetell
C 47	16.8	84.0	348706	1	EX640445	EX640445 Bordetell
C 48	16.8	84.0	349028	1	EX640413	EX640413 Bordetell
C 49	16.8	84.0	349737	1	EX572597	EX572597 Rhodopseu
C 50	16.8	84.0	349980	6	AK127150	AK127150 Sequence
C 51	16.4	82.0	1087	8	AK107440	AK107440 Oryza sat
C 52	16.4	82.0	1175	8	AK107346	AK107346 Oryza sat
C 53	16.4	82.0	1234	8	AK108703	AK108703 Oryza sat
C 54	16.4	82.0	1309	8	AK065404	AK065404 Oryza sat
C 55	16.4	82.0	1332	8	OSHOX1	X96881 Oryza sat
C 56	16.4	82.0	1410	1	AME277083	AJ277083 Amycolato
C 57	16.4	82.0	1489	8	AK105047	AK105047 Oryza sat
C 58	16.4	82.0	1514	8	AK064941	AK064941 Oryza sat
C 59	16.4	82.0	1514	8	AK100195	AK100195 Oryza sat
C 60	16.4	82.0	1531	8	AK122182	AK122182 Oryza sat
C 61	16.4	82.0	3108	8	AF211193	AF211193 Oryza sat
C 62	16.4	82.0	3578	9	EC013000	EC013000 Homo sapi
C 63	16.4	82.0	3587	6	AX676836	AX676836 Sequence
C 64	16.4	82.0	3600	6	AR366641	AR366641 Sequence
C 65	16.4	82.0	3600	6	BD141660	BD141660 Use of GI



C 285 15.4 77.0 6972 9 HAAKTRSYV  
 286 15.4 77.0 7011 6 AR198610  
 287 15.4 77.0 7011 10 RATRBB1  
 288 15.4 77.0 7420 9 HSKCNQ2  
 289 15.4 77.0 8309 6 AR339592  
 290 15.4 77.0 8902 2 AC017749  
 291 15.4 77.0 8916 1 AE012206  
 292 15.4 77.0 9695 10 AF055477  
 293 15.4 77.0 10029 1 AE009640  
 294 15.4 77.0 10155 1 AE014511  
 295 15.4 77.0 10551 1 AE004701  
 296 15.4 77.0 10982 1 AE002429  
 297 15.4 77.0 11120 1 AE004734  
 298 15.4 77.0 11219 1 AE001948  
 299 15.4 77.0 11527 1 AE005953  
 300 15.4 77.0 11819 1 AE011801

ACC09366 Drosophila  
 BS000234 Pan trogl  
 AC005696 Homo sapi  
 AC128847 Rattus no  
 AP003458 Rattus no  
 AP006440 Homo sapi  
 AP009238 Rattus no  
 AL591107 Homo sapi  
 AL124208 Rattus no  
 AC099115 Rattus no  
 AC119546 Rattus no  
 AC117369 Rattus no  
 AC127712 Rattus no  
 AC096305 Rattus no  
 AC133708 Rattus no  
 AC130856 Rattus no  
 AC107176 Rattus no  
 AC133703 Rattus no  
 AL391115 Streptomy  
 AC095749 Rattus no  
 AL591789 Sinorhizo  
 AP005045 Streptomy  
 AB017066 Rattus no  
 AB017115 Rattus no  
 AP006578 Gloeobact  
 AP006578 Gloeobact  
 AL939111 Streptomy  
 AC095174 Rattus no  
 BX572605 Rhodosphe  
 AR430774 Rattus no  
 AX028625 Sequence  
 AX028625 Sequence  
 AX028609 Sequence  
 AX056229 Sequence  
 L27995 Brucella ab  
 AX092129 Arabidops  
 AX070030 Arabidops  
 AJ330026 Homo sapi  
 AX391540 Sequence  
 AX391548 Sequence  
 AX060334 Rattus no  
 AK104544 Rattus no  
 AR430774 Rattus no  
 AX028625 Sequence  
 AX028625 Sequence  
 AF156692 Neisseria  
 AK063164 Rattus no  
 AK063164 Rattus no  
 L79946 Homo sapien  
 AP038173 Homo sapi  
 AP052089 Homo sapi  
 AF358663 Brucella  
 AK101216 Rattus no  
 AK101162 Rattus no  
 M80792 Pseudomonas  
 AK130880 Homo sapi  
 AR430772 Rattus no  
 AX028621 Sequence  
 AF277626 Synthetic  
 AC014538 Drosophila  
 AK120611 Rattus no  
 Y13070 S.coelicolo  
 AK072069 Rattus no  
 AK067999 Rattus no  
 AB082951 Homo sapi  
 X63453 D.melanogas  
 BX649015 Homo sapi  
 AB037724 Homo sapi  
 BX537556 Homo sapi  
 AX405958 Sequence  
 X62666 Pseudomonas  
 AY090863 Homo sapi  
 AR304059 Sequence

ALIGNMENTS

RESULT 1  
 MBO298910  
 LOCUS MBO298910 335 bp DNA linear BCT 23-OCT-2001  
 DEFINITION Mycobacterium bovis VNTR, strain 1503/96.  
 ACCESSION AJ298910  
 VERSION AJ298910.1 GI:16416365  
 KEYWORDS repetitive element; tandem repeat; VNTR.  
 SOURCE Mycobacterium bovis  
 ORGANISM Mycobacterium bovis  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
 tuberculosis complex.  
 REFERENCE 1  
 AUTHORS Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.  
 TITLE Fluorescent VNTR (fVNTR) for improved high throughput molecular  
 typing of Mycobacterium bovis  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 335)  
 AUTHORS Llewellyn-Hughes, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial  
 Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,  
 Surrey, KT15 3NB, UNITED KINGDOM

FEATURES

source  
 1. .335  
 /organism="Mycobacterium bovis"  
 /mol\_type="genomic DNA"  
 /strain="1503/96"  
 /db\_xref="taxon:1765"  
 repeat\_region 73..127  
 /rpt\_family="VNTR F2.1"  
 repeat\_region 128..206  
 /rpt\_type="TANDEM"  
 repeat\_region 207..292  
 /rpt\_family="VNTR F2.1"  
 /rpt\_type="TANDEM"

ORIGIN

Query Match 100.0%; Score 20; DB 1; Length 335;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 GCGCCGCCACCTCGACTT 20  
 Db 37 GCGCCGCCACCTCGACTT 56

RESULT 2

MBO298911  
 LOCUS MBO298911 335 bp DNA linear BCT 23-OCT-2001

```

DEFINITION   Mycobacterium bovis VNTR, strain 4705.
ACCESSION    AJ298911
VERSION      GI:16416366
KEYWORDS     repetitive element; tandem repeat; VNTR.
SOURCE       Mycobacterium bovis
ORGANISM     Mycobacterium bovis
              Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1
AUTHORS      Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE        Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
JOURNAL      Unpublished
AUTHORS      Llewellyn-Hughes, S.
TITLE        (bases 1 to 335)
JOURNAL      Direct Submission
              Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
FEATURES     source
              1..335
              /organism="Mycobacterium bovis"
              /mol_type="genomic DNA"
              /strain="4705"
              /db_xref="taxon:1765"
              73..127
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM
              128..206
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM
              207..292
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM
ORIGIN
Query Match      100.0%; Score 20; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCGCGCACCACTCGACTT 20
Db 37 GCGCGCACCACTCGACTT 56
RESULT 3
MMI298907
LOCUS         Mycobacterium microti VNTR, strain 4777/97.
DEFINITION   Mycobacterium microti VNTR, strain 4777/97.
ACCESSION    AJ298907
VERSION      GI:16416362
KEYWORDS     repetitive element; tandem repeat; VNTR.
SOURCE       Mycobacterium microti
ORGANISM     Mycobacterium microti
              Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1
AUTHORS      Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE        Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
JOURNAL      Unpublished
AUTHORS      Llewellyn-Hughes, S.
TITLE        (bases 1 to 390)
JOURNAL      Direct Submission
              Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
FEATURES     source
              1..390
              /organism="Mycobacterium microti"
              /mol_type="genomic DNA"
              /strain="4705"
              /db_xref="taxon:1806"
              73..127
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              128..182
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              183..261
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
ORIGIN
Query Match      100.0%; Score 20; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCGCGCACCACTCGACTT 20
Db 37 GCGCGCACCACTCGACTT 56
RESULT 5
MMI298905
LOCUS         Mycobacterium microti VNTR, strain 4700/97.
DEFINITION   Mycobacterium microti VNTR, strain 4700/97.
ACCESSION    AJ298909
VERSION      GI:16416364
KEYWORDS     repetitive element; tandem repeat; VNTR.
SOURCE       Mycobacterium microti
ORGANISM     Mycobacterium microti
              Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1
AUTHORS      Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE        Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
JOURNAL      Unpublished
AUTHORS      Llewellyn-Hughes, S.
TITLE        (bases 1 to 390)
JOURNAL      Direct Submission
              Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
FEATURES     source
              1..390
              /organism="Mycobacterium microti"
              /mol_type="genomic DNA"
              /strain="4700/97"
              /db_xref="taxon:1806"
              73..127
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              128..182
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              183..261
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
ORIGIN
Query Match      100.0%; Score 20; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCGCGCACCACTCGACTT 20
Db 37 GCGCGCACCACTCGACTT 56
RESULT 5
MMI298905

```

```

/strain="4777/97"
/db_xref="taxon:1806"
73..127
/rpt_family="VNTR F2.2"
/rpt_type=TANDEM
128..182
/rpt_family="VNTR F2.2"
/rpt_type=TANDEM
183..261
/rpt_family="VNTR F2.2"
/rpt_type=TANDEM
ORIGIN
Query Match      100.0%; Score 20; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCGCGCACCACTCGACTT 20
Db 37 GCGCGCACCACTCGACTT 56
RESULT 4
MMI298909
LOCUS         Mycobacterium microti VNTR, strain 4700/97.
DEFINITION   Mycobacterium microti VNTR, strain 4700/97.
ACCESSION    AJ298909
VERSION      GI:16416364
KEYWORDS     repetitive element; tandem repeat; VNTR.
SOURCE       Mycobacterium microti
ORGANISM     Mycobacterium microti
              Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1
AUTHORS      Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE        Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
JOURNAL      Unpublished
AUTHORS      Llewellyn-Hughes, S.
TITLE        (bases 1 to 390)
JOURNAL      Direct Submission
              Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
FEATURES     source
              1..390
              /organism="Mycobacterium microti"
              /mol_type="genomic DNA"
              /strain="4700/97"
              /db_xref="taxon:1806"
              73..127
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              128..182
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              183..261
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
ORIGIN
Query Match      100.0%; Score 20; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCGCGCACCACTCGACTT 20
Db 37 GCGCGCACCACTCGACTT 56
RESULT 5
MMI298905

```

```

LOCUS      MBO298905                414 bp    DNA          linear      BCT 23-OCT-2001
DEFINITION Mycobacterium bovis VNTR, strain AN 5.
ACCESSION  AJ298905
VERSION    AJ298905.1 GI:16416360
KEYWORDS   repetitive element; tandem repeat; VNTR.
SOURCE     Mycobacterium bovis
ORGANISM   Mycobacterium bovis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
            tuberculosis complex.
REFERENCE  1
AUTHORS    Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE      Fluorescent VNTR (fVNTR) for improved high throughput molecular
            typing of Mycobacterium bovis
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 414)
AUTHORS    Llewellyn-Hughes, S.
TITLE      Direct Submission
JOURNAL    Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
            Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
            Surrey, KT15 3NB, UNITED KINGDOM
FEATURES   Location/Qualifiers
            source             1..414
                                /organism="Mycobacterium bovis"
                                /mol_type="genomic DNA"
                                /strain="AN 5"
                                /db_xref="taxon:1765"
            repeat_region      73..127
                                /rpt_family="VNTR F2.1"
                                /rpt_type=TANDEM
            repeat_region      128..206
                                /rpt_family="VNTR F2.1"
                                /rpt_type=TANDEM
            repeat_region      207..292
                                /rpt_family="VNTR F2.1"
                                /rpt_type=TANDEM
ORIGIN
Query Match      100.0%; Score 20; DB 1; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1  GCGCGCACCACTCGACTT 20
      |||||
DB   37  GCGCGCACCACTCGACTT 56

RESULT 7
MBO298908
LOCUS      MBO298908                493 bp    DNA          linear      BCT 23-OCT-2001
DEFINITION Mycobacterium bovis VNTR, strain 2259-93.
ACCESSION  AJ298908
VERSION    AJ298908.1 GI:16416363
KEYWORDS   repetitive element; tandem repeat; VNTR.
SOURCE     Mycobacterium bovis
ORGANISM   Mycobacterium bovis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
            tuberculosis complex.
REFERENCE  1
AUTHORS    Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE      Fluorescent VNTR (fVNTR) for improved high throughput molecular
            typing of Mycobacterium bovis
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 493)
AUTHORS    Llewellyn-Hughes, S.
TITLE      Direct Submission
JOURNAL    Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
            Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
            Surrey, KT15 3NB, UNITED KINGDOM
FEATURES   Location/Qualifiers
            source             1..493
                                /organism="Mycobacterium bovis"
                                /mol_type="genomic DNA"
                                /strain="2259-93"
                                /db_xref="taxon:1765"
            repeat_region      73..127
                                /rpt_family="VNTR F4.1"
                                /rpt_type=TANDEM
            repeat_region      128..206
                                /rpt_family="VNTR F4.1"
                                /rpt_type=TANDEM
            repeat_region      207..285
                                /rpt_family="VNTR F4.1"
                                /rpt_type=TANDEM
ORIGIN
Query Match      100.0%; Score 20; DB 1; Length 493;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1  GCGCGCACCACTCGACTT 20
      |||||
DB   37  GCGCGCACCACTCGACTT 56

RESULT 8
MTU298906
LOCUS      MTU298906                469 bp    DNA          linear      BCT 23-OCT-2001
DEFINITION Mycobacterium tuberculosis VNTR, strain H37Rv.
ACCESSION  AJ298906
VERSION    AJ298906.1 GI:16416361
KEYWORDS   repetitive element; tandem repeat; VNTR.
SOURCE     Mycobacterium tuberculosis
ORGANISM   Mycobacterium tuberculosis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
            tuberculosis complex.
REFERENCE  1
AUTHORS    Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE      Fluorescent VNTR (fVNTR) for improved high throughput molecular
            typing of Mycobacterium bovis
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 469)
AUTHORS    Llewellyn-Hughes, S.
TITLE      Direct Submission
JOURNAL    Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
            Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
            Surrey, KT15 3NB, UNITED KINGDOM
FEATURES   Location/Qualifiers
            source             1..469
                                /organism="Mycobacterium tuberculosis"

```



```

/gene="WT2995"
/Note="similar to GP:5114183; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ribonuclease III"
/protein_id="AAK47322.1"
/db_xref="GI:13882776"
/translation="MISRSQPLDALGVDPDELLSALHRSYAYENGGLPTNERLE
FRGDAVLGTTTDLFRHPRDSECDLAKFASVNTQALADVARRICAEGLGVHVL
CLGANTGADKSSITLADGMSLGAITYLOHMEKAREVILRLFGPLDIAAPTIGAGL
DKWTSLOBLTAARGLGAPSVLVTSTGPDHDKFTFAVVVMDSYGVSGVGRSKAEQK
AAAAAKALEVLDNAMPKTSK"
complement(7924..8511)
/gene="WT2996"
complement(7924..8511)
/gene="WT2996"
/Note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK47323.1"
/db_xref="GI:13882777"
/translation="MAHQGPTAQHVASPMVDIARLGRPGAMFELHDTVHSPARI
GLETAIDQALLDLIRVESVSEGLVTGTVAATVGECAKCLSPVGRGVQVALTEL
FAYPDSATDEETEDVGRVDETDLEQPTIDAVGLELPSPVCRDPCGLCPQCGV
PLAEPGCHRHQIDPRWAKLVEMLGPSDILGR"
complement(8598..9341)
/gene="WT2997"
complement(8598..9341)
/gene="WT2997"
/Note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK47324.1"
/db_xref="GI:13882778"
/translation="MAVYRVEALDELSAIVEEARGVPMTAGCVPRGDVLELIDDIK
DAIPGELDDAODVLDARDMLQDAKTHADSMVSSATTAEKSLNHARTEDRLSDAK
AQADRMVSEARQSRMVMADAEARIAATAKREYASVRAQACDRLIENGNIY
AKAVQBGKEQORLSQNEVAAAESTRVLDTAHAEADRLGECDIYVDNKLAEFE
BFLNGTUSVGRGRHQRLKTAAGTHDIYVR"
9568..10359
/gene="WT2998"
9568..10359
/gene="WT2998"
/Note="similar to GB:U08223 PID:473601; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="AAK47325.1"
/db_xref="GI:13882779"
/translation="MSMLARHPYGVGVNGHSDSSGDAKQAKAPTLIYFHAGGTAK
DYAFSREFSDVRIAYVQVPGQHDRGLPLESIPLADEIFAMRPSARIDDPVAF
Query Match 100.0%; Score 20; DB 1; Length 10894;
Best Local Similarity 100.0%; Pred.No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGGACACCTCGACTT 20
|||||
Db 6811 GCGCGGACACCTCGACTT 6830

RESULT 9
BX248344 318050 bp DNA linear BCT 11-JUN-2003
LOCUS
DEFINITION
Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment
11/14.
ACCESSION
BX248344 BX248333
VERSION
BX248344.1 GI:31619628
KEYWORDS
complete genome.

```

```

SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL

REMARK
REFERENCE

AUTHORS
TITLE
JOURNAL

FEATURES
source

gene
CDS

gene
CDS

gene
CDS

```

```

Mycobacterium bovis subsp. bovis AF2122/97
Mycobacterium bovis subsp. bovis AF2122/97
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1
Garnier,T., Eglmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
Pryor,M., Duthoy,S., Grondin,S., Iacoiu,C., Mensepe,C., Simon,S.,
Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,
Wheeler,P.R., Parkhill,J., Barrell,B.G., Cole,S.T., Gordon,S.V. and
Hewison,G.
The complete genome sequence of Mycobacterium bovis
Online Publication
PNAS 10.1073/pnas.1130426100 ( Microbiology )
2 (bases 1 to 318050)
Garnier,T.
Direct Submission
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
Bacterienne Institut Pasteur 28,rue du Dr Roux 75724 PARIS cedex
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
Laboratories Agency Weybridge, Woodham Lane, New Haw,Addlestone,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton,Cambridge CB10 1SA, UK. Pr4 Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
rue du Docteur Roux, 75724 Paris Cedex 15, France
Location/Qualifiers
1. 318050
/organism="Mycobacterium bovis subsp. bovis AF2122/97"
/mol_type="Genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"
complement(99..875)
/locus_tag="Mb2882c"
complement(99..875)
/locus_tag="Mb2882c"
/EC_number="1.-.-.-"
/Note="Mb2882c", len: 258 aa. Equivalent to Rv2857c,
len: 258 aa. from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 258 aa overlap). Probable short-chain
dehydrogenase/reductase (EC 1.-.-.-), highly similar to
various dehydrogenases e.g. O88068|SCI35.33c PROBABLE
DEHYDROGENASE (SDR FAMILY) from Streptomyces coelicolor
(260 aa), FASTA scores: opt: 1208, E(): 2e-68, (72.35%
identity in 253 aa overlap); Q9I376|PA1649 from
Pseudomonas aeruginosa PROBABLE SHORT-CHAIN DEHYDROGENASE
(253 aa), FASTA scores: opt: 569, E(): 2.1e-28, (39.2%
identity in 255 aa overlap); Q9EX74|MLHA SCR-LIKE ENZYME
from Rhodococcus erythropolis (246 aa), FASTA scores: opt:
567, E(): 2.8e-28, (41.1% identity in 248 aa overlap);
etc. Also similar to many Mycobacterium tuberculosis
dehydrogenases e.g. FABG3|Rv2002|MT2058|MTCY39.16c
PUTATIVE OXIDOREDUCTASE (260 aa), FASTA score: (38.3%
identity in 248 aa overlap). BELONGS TO THE SHORT-CHAIN
DEHYDROGENASES/REDUCTASES (SDR) FAMILY..
/codon_start=1
/transl_table=11
/product="PROBABLE SHORT-CHAIN TYPE
DEHYDROGENASE/REDUCTASE"
/protein_id="CAD96569.1"
/db_xref="GI:31619629"
/translation="MMDLSRLAGRVAVITGGSGIGLAAGRRMAEGATIVGVGDV
RAGAAADELSGLVPTDCDDAVNGLFDGAETVGRIDIAFNAGISPPBDNLLEN
TELAAMQVQDVNLKSVLLCCRAALHMHVLAGKSVNTAFVAVNGATSIKSTAS
KGVGLAMSRELGVQFARQIRVNALCPGPVNTPLQELPAKPERAARMVHPLGRF
APDEIAAAVAFLASDASIFATSTFLVDGGISSAYVTEL"
complement(872..2239)
/gene="aldC"
/locus_tag="Mb2883c"
complement(872..2239)
/gene="aldC"
/locus_tag="Mb2883c"

```





```
XX Mycobacterium tuberculosis H37Rv
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium tuberculosis complex; Mycobacterium tuberculosis.
XX
XX [1]
RX MEDLINE; 98295987.
RX PUBMED; 9634230.
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry III C.E., Tekala F., Badcock K.,
RA Bigham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K.,
RA Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K.,
RA Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J.,
RA Quail A.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S.,
RA Squares S., Squires R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete
RL genome sequence";
RL Nature 393:537-544 (1998).
XX
XX [2]
RX PUBMED; 12368430.
RA Camus J.C., Pryor M.J., Medigue C., Cole S.T.;
RA "Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv";
RL Microbiology 148:2967-2973 (2002).
XX
XX [3]
RX 1-348676
RA Parkhill J.;
RT ;
RL Submitted (11-JUN-1998) to the EMBL/GenBank/DBJ databases.
RL Submitted on behalf of the Mycobacterium tuberculosis sequencing and
RL mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton,
RL Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut
RL Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail:
RL parkhill@sanger.ac.uk
XX
CC Notes:
CC Details of M. tuberculosis sequencing at the Sanger Centre
CC are available on the World Wide Web
CC (URL, http://www.sanger.ac.uk/Projects/M\_tuberculosis/)
XX
XX Key
XX Location/Qualifiers
XX
XX source
XX 1. 348676
XX /db_xref="taxon:8332"
XX /mol_type="genomic DNA"
XX /organism="Mycobacterium tuberculosis H37Rv"
XX /strain="H37Rv"
XX complement(148.183)
XX repeat_unit
XX /note="36 bp direct repeat, 36 out of 36 bp identical to
XX sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
XX complement(148.4539)
XX repeat_region
XX /note="4392 bp direct repeat region"
XX complement(223.257)
XX repeat_unit
XX /note="36 bp direct repeat, 35 out of 36 bp identical to
XX sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
XX complement(298.333)
XX repeat_unit
XX /note="36 bp direct repeat, 35 out of 36 bp identical to
XX sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
XX complement(374.409)
XX repeat_unit
XX /note="36 bp direct repeat, 36 out of 36 bp identical to
XX sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
XX complement(447.482)
XX repeat_unit
XX /note="36 bp direct repeat, 35 out of 36 bp identical to
XX sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
XX complement(519.554)
XX repeat_unit
XX /note="36 bp direct repeat, 36 out of 36 bp identical to
XX sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
XX complement(590.625)
XX repeat_unit
XX /note="36 bp direct repeat, 36 out of 36 bp identical to
XX sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
XX complement(664.699)
XX repeat_unit
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(740.775)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(811.846)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(884.919)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(958.993)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(1031.1086)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(1104.1139)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(1176.1211)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(1248.1283)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(1322.1357)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(1396.1431)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(1467.1486)
FT /note="20 bp partial direct repeat, CCCCAGAGGGGACGGAAC,
FT of sequence GTCGTACAGCCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(1486.2860)
FT /note="IS6110-11, len: 1375 bp. Insertion sequence IS6110."
FT /insertion_seq="IS6110-11"
FT complement(1529.2467)
FT /note="RV2814c. (MTCY16B7.29), len: 312 aa. Probable
FT transposase, highly similar to others e.g.
FT P97137|RV0796|MTV042.06 PUTATIVE TRANSPOSASE FOR INSERTION
FT SEQUENCE ELEMENT IS986/IS6110 from Mycobacterium
FT tuberculosis (328 aa), FASTA scores: opt: 2103, E():
FT 6.1e-132, (100.0% identity in 312 aa overlap); etc. Start
FT unlikely."
FT /transl_table=11
FT /function="REQUIRED FOR THE TRANSPOSITION OF THE INSERTION
FT ELEMENT IS6110."
FT /locus_tag="RV2814c"
FT /product="PROBABLE TRANSPOSASE"
FT /protein_id="CAB03675.1"
FT /translation="LITRFDIADHGHGPGDLRWGVESICTQLTELGVPIAPSTYYDHI
FT INPSSRELRLDGLKEHISRVHAANYGVYGVARKVLTILNREGIVARCTVERLMTKL
FT LSGTTRGKARTTIADPATAPADLVQRFPAPNRLWVADLTYSWAGAYVAFVT
FT DAYARRILGWRVASTMAYSLDLALAEQIWRKQGVLDKDVHHTDORGSQYIRFS
FT ERLAAGQPSVAGSSYMDALAEQIWRKQGVLDKDVHHTDORGSQYIRFS
FT NHRRLYQGVDPVPELEAAYVQQRPAAG"
FT complement(2464.2790)
FT /note="RV2815c. (MTCY16B7.28), len: 108 aa. Probable
FT transposase, identical from aa 51 with P19772|YIA2_MYCTU
FT PUTATIVE TRANSPOSASE (INSERTION ELEMENT IS986) from
FT Mycobacterium tuberculosis (59 aa), FASTA scores: opt: 365,
FT E(): 1.1e-19, (96.6% identity in 59 aa overlap); and other
FT transposases."
FT /transl_table=11
FT /function="REQUIRED FOR THE TRANSPOSITION OF THE INSERTION
FT ELEMENT IS6110."
FT /locus_tag="RV2815c"
FT /product="PROBABLE TRANSPOSASE"
FT /protein_id="CAB03633.1"
FT /translation="NSGSSRRYPPELRERAVRMVAEIRGQDSEWAAI SEVAILGVG
FT CAETVRKWRVQAQVDAGARPQTITTESEALKELRDRAELRANILKTASAFFAELD
```

FT repeat\_unit  
FT complement(2845. .2860)  
FT /note="36 bp partial direct repeat, GTGCTCAGACCCAAAA, of  
FT sequence GTGCTCAGACCCAAAAACCCGAGAGGGGACGGAAC"  
FT complement(2901. .2936)  
FT /note="36 bp direct repeat, 36 out of 36 bp identical to  
FT sequence GTGCTCAGACCCAAAAACCCGAGAGGGGACGGAAC"  
FT complement(2976. .3011)  
FT /note="36 bp direct repeat, 36 out of 36 bp identical to  
FT sequence GTGCTCAGACCCAAAAACCCGAGAGGGGACGGAAC"  
FT complement(3049. .3084)  
FT /note="36 bp direct repeat, 36 out of 36 bp identical to  
FT sequence GTGCTCAGACCCAAAAACCCGAGAGGGGACGGAAC"  
FT complement(3121. .3156)  
FT /note="36 bp direct repeat, 36 out of 36 bp identical to  
FT sequence GTGCTCAGACCCAAAAACCCGAGAGGGGACGGAAC"  
FT complement(3193. .3228)  
FT /note="36 bp direct repeat, 36 out of 36 bp identical to  
FT sequence GTGCTCAGACCCAAAAACCCGAGAGGGGACGGAAC"  
FT complement(3266. .3301)  
FT /note="36 bp direct repeat, 36 out of 36 bp identical to  
FT sequence GTGCTCAGACCCAAAAACCCGAGAGGGGACGGAAC"  
FT complement(3338. .3373)  
FT /note="36 bp direct repeat, 36 out of 36 bp identical to  
FT sequence GTGCTCAGACCCAAAAACCCGAGAGGGGACGGAAC"  
FT complement(3399. .3434)  
FT /note="36 bp direct repeat, 36 out of 36 bp identical to  
FT sequence GTGCTCAGACCCAAAAACCCGAGAGGGGACGGAAC"  
FT complement(3476. .3511)  
FT /note="36 bp direct repeat, 36 out of 36 bp identical to  
FT sequence GTGCTCAGACCCAAAAACCCGAGAGGGGACGGAAC"  
FT complement(3548. .3583)  
FT /note="36 bp direct repeat, 36 out of 36 bp identical to  
FT sequence GTGCTCAGACCCAAAAACCCGAGAGGGGACGGAAC"  
FT complement(3624. .3659)  
FT /note="36 bp direct repeat, 36 out of 36 bp identical to  
FT sequence GTGCTCAGACCCAAAAACCCGAGAGGGGACGGAAC"  
FT complement(3701. .3736)  
FT /note="36 bp direct repeat, 36 out of 36 bp identical to  
FT sequence GTGCTCAGACCCAAAAACCCGAGAGGGGACGGAAC"  
FT complement(3774. .3809)

Query Match 100.0%; Score 20; DB 15; Length 348676;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGCACCCTCGACTT 20  
DB 120398 GCGCGCACCCTCGACTT 120417

RESULT 11  
MSGB1723CS/c 38477 bp DNA linear BCT 17-DEC-2001  
LOCUS Mycobacterium leprae cosmid B1723 DNA sequence.  
DEFINITION L78825  
ACCESSION L78825  
VERSION L78825.1 GI:1377772  
KEYWORDS  
SOURCE Mycobacterium leprae  
ORGANISM  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
REFERENCE  
1 (sites)  
Eiglmeyer, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.  
Use of an ordered cosmid library to deduce the genomic organization  
of Mycobacterium leprae  
Mol. Microbiol. 7 (2), 197-206 (1993)  
JOURNAL 93188700  
MEDLINE 8446027  
PUBMED  
REFERENCE  
2 (bases 1 to 38477)  
Smith, D.R., Richterich, P., Rubenfield, M., Rice, P.W., Butler, C.,  
Lee, H.M., Kirt, S., Gundersen, K., Abendschan, K., Xu, Q., Chung, M.,  
Deloughery, C., Aldredge, T., Maher, J., Lundstrom, R., Tulig, C.,

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Falls, K., Imrich, J., Torrey, D., Engelstein, M., Breton, G., Madan, D.,  
Nietupski, R., Seitz, B., Mao, J.I. et al.  
Multiplex sequencing of 1.5 Mb of the Mycobacterium leprae genome  
Genome Res. 7 (8), 802-819 (1997)  
97413161  
9267804  
Original source text: Mycobacterium leprae (clone: cosmid B1723)  
(tissue library: Loris 6) DNA.  
This sequence was generated by the Genome Sequencing Center at  
Genome Therapeutics Corporation (Collaborative Research Division),  
100  
Beaver St., Waltham, MA, 02154. Please contact Doug Smith  
(smith@eric.com) for further information. The sequence represents  
the  
insert of a Loris 6 cosmid clone from a mapped set of clones  
constructed from M. leprae genomic DNA isolated from armadillo  
liver  
[3]. The sequence may not represent the entire cloned insert of  
the  
cosmid if an overlapping region was previously sequenced from  
another  
clone. Coding sequences larger than 60 amino acids were predicted  
on  
the basis of codon usage and homology information. An attempt was  
made  
to locate the most probable start site based on codon usage,  
homology,  
the presence of a Shine-Dalgarno sequence, or overlapping orf that  
suggested translational coupling. It is possible that the actual  
start  
site differs from the one selected.

FEATURES  
source

Location/Qualifiers  
1. .38477  
/organism="Mycobacterium leprae"  
/mol\_type="genomic DNA"  
/specific\_host="Dasypus novemcinctus"  
/db\_xref="taxon:1769"  
/clone="cosmid B1723"  
/tissue\_lib="Loris 6"  
/note="The liver of the armadillo was used to isolate the  
Mycobacterium leprae."

ORIGIN

Query Match 92.0%; Score 18.4; DB 1; Length 38477;  
Best Local Similarity 95.0%; Pred. No. 5.3e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGCGCACCCTCGACTT 20  
DB 23828 GCGCGCACCCTCGACTT 23809

RESULT 12  
MLCB1243/c 42926 bp DNA linear BCT 27-AUG-1999  
LOCUS Mycobacterium leprae cosmid B1243.  
DEFINITION AL023635  
ACCESSION AL023635  
VERSION AL023635.1 GI:3150214  
KEYWORDS  
accD6; aceB; Acetyl/Propionyl CoA Carboxylase; acpM; acyl carrier  
protein; ATP-dependent DNA helicase; beta-ketoacyl-ACP synthase;  
formamidopyrimidine-dna glycosylase; fpg; kasA; kasB; kdsB;  
lipopolysaccharide core biosynthesis protein; malonyl CoA-acyl  
carrier protein transacylase; mutM; oxidoreductase; pyruvate  
dehydrogenase e1 component; recG; ribonuclease III; RLEP; rnc.  
SOURCE Mycobacterium leprae  
ORGANISM  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
REFERENCE  
1 (bases 1 to 42926)  
Eiglmeyer, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.  
Use of an ordered cosmid library to deduce the genomic organization  
of Mycobacterium leprae  
Mol. Microbiol. 7 (2), 197-206 (1993)  
JOURNAL



/gene="MLCB1243.03c"  
 /note="P500062 Aldc/keto reductase family signature 2"  
 complement(3448. .4254)  
 /note="MLCB1243.04c, probable oxidoreductase pseudogene,  
 similar to e.g. 2DKG CORSP 2,5-diketo-d-gluconic acid  
 reductase and to downstream ORF MLCB1243.03c"  
 /pseudo  
 /codon\_start=1  
 /transl\_table=11  
 /complement(3450. .3453)  
 /note="possible RBS upstream of MLCB1243.03c"  
 4318. .5383  
 /note="MLCB1243.05, probable pseudogene, equivalent to M.  
 tuberculosis hypothetical lipase/esterase TR:P95125  
 (EMBL:Z83018) MTCV349.17 (376 aa)"  
 /pseudo  
 /codon\_start=1  
 /transl\_table=11  
 5457. .5460  
 /note="possible RBS upstream of MLCB1243.06"  
 5472. .6248  
 /gene="MLCB1243.06"  
 5472. .6248  
 /note="MLCB1243.06"  
 /note="MLCB1243.06, possible transmembrane protein, len:  
 258 aa; equivalent to M. tuberculosis hypothetical  
 membrane protein TR:O33272 (EMBL:Z83018) MTCV349.18 (255  
 aa). fasta scores; Opt: 1097 z-score: 1595.4 E(): 0, 65.5%  
 identity in 258 aa overlap"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative transmembrane protein"  
 /protein\_id="CAA19187.1"  
 /db\_xref="GI:3150218"  
 /db\_xref="GOA:O69463"  
 Query Match 92.0%; Score 18.4; DB 1; Length 42926;  
 Best Local Similarity 95.0%; Pred. No. 5.3e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GCGCGCACCACTCGACTT 20  
 |||||  
 Db 15698 GCGCGCACCACTCTACTT 15679  
 RESULT 13  
 MLEPRTN7  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mycobacterium leprae  
 Mycobacterium leprae  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 Cole, S.T., Eigmler, K., Parkhill, J., James, K.D., Thomson, N.R.,  
 Wheeler, P.R., Honore, N., Ganier, T., Churcher, C., Harris, D.,  
 Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R.,  
 Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A.,  
 Hamlin, N., Holroyd, S., Hornsby, T., Jags, K., Lacroix, C.,  
 Maclean, J., Moule, S., Murphy, L., Oliver, Quail, M.A.,  
 Rajandream, M.A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S.,  
 Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K.,  
 Taylor, K., Whitehead, S., Woodward, J.R. and Barrall, B.G.  
 Massive gene decay in the leprosy bacillus  
 Nature 409 (6823), 1007-1011 (2001)  
 21128732  
 MEDLINE  
 PUBLISHED  
 11234002  
 2 (bases 1 to 348950)  
 REFERENCE  
 JOURNAL  
 TITLE  
 Parkhill, J.  
 Direct Submission

JOURNAL  
 COMMENT  
 Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium  
 leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome  
 Campus, Hinxton, Cambridge, CB10 1SA, UK Unite de Genetique  
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,  
 75724, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk  
 Notes:  
 Details of M. leprae sequencing at the Sanger Centre are available  
 from http://www.sanger.ac.uk/Projects/M\_leprae/ A relational  
 database containing the M. leprae sequences is available from  
 http://genolist.pasteur.fr/Leprae/  
 Location/Qualifiers  
 FEATURES  
 source  
 1. .348950  
 /organism="Mycobacterium leprae"  
 /mol\_type="genomic DNA"  
 /strain="TN"  
 /db\_xref="taxon:1769"  
 95. .794  
 /note="Dispersed repeat, RLEP, copy 27"  
 complement(1041. .1889)  
 /gene="fpg"  
 /note="synonym: MLI1658"  
 complement(1041. .1889)  
 /gene="fpg"  
 /note="Highly similar to many formamidopyrimidine-DNA  
 glycosylases (EC 3.2.2.23) thought to be involved in DNA  
 repair after exposure to alkylating agents e.g.  
 Escherichia coli SW:FPG\_ECOLI (P05523) (215 aa); BlastP  
 Expect 8.7 and Mycobacterium tuberculosis SW:FPG\_MTCU  
 (Q10959) (289 aa); Fasta score E(): 0, 80.6% identity in  
 289 aa overlap. Contains Pfam match to entry PF01149  
 Fapy DNA glyco, Formamidopyrimidine-DNA glycosylase."  
 /codon\_start=1  
 /transl\_table=11  
 /product="formamidopyrimidine-DNA glycosylase"  
 /protein\_id="CAC30611.1"  
 /db\_xref="GI:13093429"  
 /db\_xref="GOA:O69470"  
 /db\_xref="SWISS-PROT:O69470"  
 /translation="MEPELVSVRGLDYIVGKTTAVRVHPRVRRHVAGPTDL  
 TNKLGTRINGDRGKYIWFLLDDIALVHLGMSGMLCTVPRVDHVRISALFDD  
 GTVNFDTQRTLGWLLADMTVDGSLVLPVVAHLARDPDPFRFVEAVKVLKCKHS  
 ELKRQLDQQTSGIGNTYADEALMRAEVHGAIAATLRQLAAVLDAAADVMDKSL  
 AKGTSFSLYVNVNGESGYFDRSLDAYGREGGCRGAVMREKFMNRSFYCPRC  
 QPRPRR"  
 complement(1059. .1889)  
 /gene="fpg"  
 /note="Pfam match to entry PF01149 Fapy DNA glyco,  
 Formamidopyrimidine-DNA glycosylase, score 531.10, E-value  
 7.8e-156"  
 complement(1944. .2660)  
 /gene="rnc"  
 /note="synonym: MLI1659"  
 complement(1944. .2660)  
 /gene="rnc"  
 /note="Highly similar to many ribonucleases involved in  
 the digestion of double stranded RNA e.g. Escherichia coli  
 SW:RNC\_ECOLI (P05737) (226 aa); Fasta score E(): 1e-24,  
 41.8% identity in 213 aa overlap and Mycobacterium  
 tuberculosis RV2925C SW:RNC\_MTCU (Q10962) (240 aa); Fasta  
 score E(): 0, 86.8% identity in 235 aa overlap. Contains  
 Pfam match to entry PF00035 dsrm, Double-stranded RNA  
 binding motif. Contains Pfam match to entry PF00636  
 Ribonuclease 3, RNase3 domain. . Contains FS00517  
 Ribonuclease III family signature."  
 /codon\_start=1  
 /transl\_table=11  
 /product="RNase III"  
 /protein\_id="CAC30612.1"  
 /db\_xref="GI:13093430"  
 /db\_xref="GOA:O69469"  
 /db\_xref="SWISS-PROT:O69469"  
 /translation="MTQPRQALDAGFVLDLPDELALTHRSYAYEHGGLPTNEBLE  
 FLGDAVLSLITITDELFRHPRSEGDGLAKLRASVNTQALAYARNLSDGGLGYLLL